December 1997 DCIEM No. 97-R-66

AN EVALUATION OF WORKLOAD MODEL PREDICTIONS IN A HELICOPTER ENVIRONMENT

B. Cain

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
North York, Ontario
Canada M3M 3B9

DEPARTMENT OF NATIONAL DEFENCE - CANADA

Executive Summary

This report compares the empirical workload results obtained during a field exercise involving four CH-136 Kiowa crews with results predicted from a task network simulation of the exercise.

None of the correlation coefficients is outstanding and only a few of the analytical measures explained greater than 50% of the variance in the empirical workload scores. The correlations between the pilots' empirical workload scores and the analytical workload values were generally greater than that found for the observers, possibly a result of the smaller number of subjects in the observer group or perhaps reflecting greater attention focused on the workload of the pilot by the modelling community. The variation of the workload measures within each flight was substantial, suggesting individual differences between subjects as well as differences in the details of each mission played significant roles in determining the perception of workload rated by the subjects.

Of the overall workload measures, the simpler measures were found to capture the greatest portion of the empirical workload variance although these measures provide little detail in what is actually causing the overload and at best only capture 50% of the workload variance. While the multi-dimensional workload measures may provide greater detail about what is causing high workloads, they do not seem to be capturing a great deal of the workload variance to begin with.

Although the correlations found in this study are low and only half the workload variance was captured, the models may still be useful. The methods currently in use, however, are still immature and require that the user have detailed knowledge of their capabilities and limitations. Considerably more research followed by thoughtful development is necessary before this analysis approach should be promoted as a robust and proven tool to the design community.

Table of Contents

Executive Summary i

Introduction 1

Method 1

Empirical Investigation 1

Analytical Workload Assessment 3

Analysis Procedure 9

Results 13

Discussion 19

Uni-dimensional Workload Measures 20

Multi-dimensional Workload Models 22

Multi-dimensional workload components 24

Conclusion 31

References 35

Appendix A. A1

Appendix B. B1

Introduction

A model producing an accurate assessment and prediction of operator cognitive workload would be a useful tool for designers of equipment or systems in which the human is the primary decision-maker. The aircraft industry is particularly interested in this topic because of the increasing technological complexity of aircraft as well as the increasing demands on the flight crew by the environments in which the aircraft is being flown. While numerous models of cognitive workload have been proposed, no single model has yet proven to be adequate in all cases. Thus, efforts continue to either refine existing models or develop new, hopefully better, models.

The purpose of this study was to compare the assessment of operator workload from several models with empirical subjective ratings in order that the validity of the models could be examined. This study is based on workload assessments made during training exercises in CH-136 Kiowa helicopters. The analytical assessments were made based on representative scenarios developed from video recordings of these flights (Hendy [1], Shaffer et al. [2]).

Method

Empirical Investigation

Several Canadian Forces flight crews were video-taped during ground-support exercises in CH-136 Kiowa helicopters. The mission for each flight was based on a similar scenario, scripted to study Kiowa workload issues in a realistic environment, and the exercise spanned several days. Four pilots and three observers were involved in the study; two of the pilots and one observer participated twice; not all flight crews had the opportunity to repeat the mission. Because of the small number of subjects, the repetitions were included in the analysis, despite the potential for biasing the results.

Each flight crew, in the presence of the investigator, reviewed the video-tape recording of their mission, scoring various aspects of workload during specific segments of

the mission. Most of the missions were reviewed immediately after the flights, however, due to scheduling constraints, two missions were reviewed the next day. These delays were not thought to result in an appreciable change in subjective workload ratings that were collected as it has been reported that ratings obtained during trials, immediately after trials, and after considerable delay but with the memory aid of video, all are highly correlated and have similar absolute values (Hart et al. [3], Haworth et al. [4]).

The flights were sufficiently similar that they could be broken into 13 common segments. Nine segments that spanned a range of workload levels were selected, analyzed and used to create task network models. The 9 segments that were modelled are:

(1) Take-off and departure;
(2) Transit to exercise area;
(3) Ingress to exercise area;
(4) Rendezvous briefing and departure;
(5) Advance to contact;
(6) Prosecution of contact;
(7) Re-deployment and re-tasking;
(8) Egress from exercise area;
(9) Rejoin and landing.

While reviewing the video-tape, each flight crew member scored a representative five minute portion of each segment, recording perceptions of their own overall workload using the method of magnitude estimation (Gopher and Braun [5]). The take-off-and-departure segment was selected as the reference flight condition against which the other segments were judged in the magnitude estimation scoring; its value was prescribed to be 100 and the subjects were instructed to score other segments relative to the reference segment value with whatever relationship they thought appropriate. At the same time, the subjects assessed their perceived demands on several aspects of workload using 21-point, doubly-anchored scales of the following factors noting the endpoint anchors used:

Factor Name	Score 0	to	Score 20
Task Complexity	Simple	-	Complex
Activity Level	Leisurely	-	Frantic
Mental Effort	Hardly at all	-	Completely absorbed
Physical Effort	Hardly at all	-	All my effort
Anxiety Level (Pucker factor)	At ease	-	Anxious
Own Performance	Best Yet	-	Off Day

These factors were presumed to be components of workload (Hendy [1]) and were selected based on similar factors considered to be important in the NASA-TLX workload model (Hart and Staveland [6]).

Analytical Workload Assessment

The video recordings of the missions were subsequently reviewed under contract, a functional decomposition performed and a typical mission time-line (an ordered series of events although the start time of each event is unspecified) was developed (Youngson et al. [7]). This mission time-line was entered into SOLE (System Operator Loading Evaluation, Anonymous [8]), a workload assessment program that incorporates the task-network analysis program SAINT (System Analysis for Integrated Networks of Tasks, Seifert [9], Wortman and Duket [10]) as well as several common workload metrics that will be described below. SOLE provides task information database management and a computer interface intended to facilitate the development and analysis of task time-lines. The time-line is decomposed into tasks for each segment of the mission time-line (a portion of the time-line with an identifiable beginning and ending) and a task network is built in SAINT. SAINT processes the task execution using data it receives from SOLE using the task characteristics to produce data for workload assessment. Tasks may occur either in parallel or in series. Concurrent tasks may be dynamically rescheduled or shed, depending upon task importance, in an attempt to mimic human behaviour for preventing or reducing operator overload conditions.

The time required to perform a task may have a stochastic component, resulting in different overlaps of tasks running in parallel during different analysis runs. The task time-line was run 100 times which was judged a sufficiently large number of iterations to obtain stable results; each run was normalized to a common duration using the average duration of several uniquely identifiable intervals within each segment to normalize each corresponding interval of each run. This normalizing process aided diagnosing which tasks were in conflict by focusing overload conditions more on tasks rather than spreading the overload condition over a broad interval which would occur if normalization was based only on total segment duration. The resulting workload measures were averaged over all runs at each second of the normalized mission. To further smooth the results, the workload values recorded each second were computed by averaging the workload values in a 60 second window centred on that second. During the first and last 30 seconds of a segment, one-half of the averaging window was truncated where data were not available.

To develop the mission time-lines, the tasks performed by the flight crew were analyzed by subject matter experts who estimated values for task characteristics (such as the completion time, the loads on cognitive resources, etc.). Similar tasks already in the SOLE database were updated for the current problem and new tasks were added as required. Each task was reviewed to determine the demands placed on the subject's visual, auditory, cognitive and psychomotor (VACP) attentional capacities as well as the likelihood that each task can be performed in conjunction with other tasks.

The four VACP component attentional demands (McCracken and Aldrich [11]) were obtained by having the subject matter experts subjectively rate each task on a 7-point, category scale. The VACP component values were subsequently transformed into interval scale values (Schuck [12]) for use in the workload calculations. The VACP scores were assigned to the appropriate W/Index attentional demand values for each task (North and Riley [13]) while the conflict coefficients for inter and intra-channel demand conflicts were adopted from Sarno and Wickens [14, 15].

The measures used to assess workload were (Youngson et al. [7]): the occurrence of unacceptable combinations of tasks; the occurrence of either unacceptable or marginally-acceptable combinations of tasks; a task conflict parameter; a time occupied parameter; the number of concurrent tasks; individual Visual, Auditory, Cognitive and Psychomotor demands; and, a modified W/Index score broken down into Within Task Demands and Across Task Interference which were subsequently added together to determine an overall W/Index score. Further descriptions of each of these measures are given below.

The assessment of the compatibility of concurrent tasks (used to calculate the occurrence of acceptable or marginally-acceptable task combinations) was calculated in a similar manner to that used at Sikorsky Aircraft (Hamilton [16]). Three sets of conflict matrices for each of the VACP components were constructed relating benchmark task activities to one-another, noting whether two concurrent tasks produced acceptable, marginally-acceptable or unacceptable demands on the operator. In instances where more than two concurrent tasks occurred, acceptability was still based on interference between pairs of tasks but all pair-wise combinations of tasks were examined; the worst case of each pair-wise comparison was assumed to be representative of the task conflict at that particular moment. The three sets of conflict matrices were: the original Sikorsky matrices; Individual matrices developed at the Defence and Civil Institute of Environmental Medicine for the Kiowa environment; and a Combined set of matrices (Schuck [12]). The Individual matrices comprised two sets of VACP conflict matrices, one set developed for the pilot and another set for the observer. The Combined set was constructed by pooling all the pilot and observer data.

The Unacceptable Task Count (UTC) represents the probability of incurring an unacceptable combination of tasks during any specific second in a mission segment based on the task interaction and the conflict matrices described above. It is calculated by averaging the occurrence of unacceptable task combinations over all runs, on a second by second basis as per the following equation:

$$UTC(t) = \sum_{run=1}^{100} \left(T(\text{unacceptable})|_t \right) / 100 \quad (1)$$

where T(unacceptable)$|_t$ is the truth value (0 for false, 1 for true) of a condition of unacceptable combination of tasks existing at time "t". The Unacceptable plus Marginal Task Count (UMTC) is a similar measure, adding the occurrence of marginal combinations of tasks to the unacceptable combinations:

$$UMTC(t) = \sum_{run=1}^{100} \left(T(\text{unacceptable or marginal})|_t \right) / 100 \quad (2)$$

The Task Conflict Parameter (TCP) represents the fraction of time that either Unacceptable or Marginally-acceptable conditions exist. It is calculated as the discrete temporal average (on a second by second basis) of the occurrence of either previously calculated Unacceptable or Marginally-acceptable conditions over a 60-second interval centred on each second of the mission segment averaged over the number of runs as described in the following equation:

$$TCP(t) = \sum_{runs=1}^{100} \left(\sum_{\tau=t-30}^{t+30} T(\text{unacceptable or marginal})|_\tau \Big/ (60*100) \right) \quad (3)$$

If Unacceptable or Marginally-acceptable conditions exist for 60% of the 60-second interval, then the Task Conflict Parameter indicates a Marginal condition; if either condition exists for 80% of the 60-second interval, then the Task Conflict Parameter indicates an Unacceptable condition. The selection of 60% and 80% of the 60-second window was arbitrary, but based on what *seemed* reasonable.

The Time Occupied Parameter (TOP) indicates (on a second by second basis) the fraction of time that an operator was engaged with at least one discrete task. Continuous tasks (such as flying the aircraft or monitoring the radio) were not included in this computation since the parameter would then always be 100%. The value is the amount of time subjects were occupied with at least one task, averaged over a sliding 60-second interval centred on each second in succession, then averaged over the number of runs as follows:

$$TOP(t) = \sum_{run=1}^{100} \sum_{\tau=t-30}^{t+30} \text{T(active discrete task)}\Big|_{\tau} \Big/ (60*100) \qquad (4)$$

where T(active discrete task)|$_\tau$ is the truth value of "Is there at least one active discrete task at time τ?". No distinction is made as to whether there are several concurrent tasks or a single task active; nor is there any attempt to assess the difficulty of the active tasks. A value of 70% of time occupied is sometimes used to indicate potential overload conditions but this "redline" value is not rigorously defined and is only a rule-of-thumb..

The Number of Concurrent Tasks (NOC), excluding continuous tasks, was recorded as an average over all runs at each second of the normalized segment of the number of tasks that were active at that time; this is an instantaneous value as a sliding time-window average was not used in this measure. The Number of Concurrent Tasks was computed as:

$$NOCT(t) = \sum_{runs=1}^{100} (\text{number of active tasks})\Big|_{t} / 100 \qquad (5)$$

This value provides a measure of subject activity level, although again, no rigorously defined value has been found to describe an overload condition. There is evidence suggesting that as the overall workload becomes increasingly complex, processing of concurrent tasks switches from a parallel process to a serial, time-sharing process: Liao and Moray [17] found that experimental evidence supported parallel processing of two concurrent tasks, however, when the number of concurrent tasks increased to four, subjects seemed to change processing strategies reverting to a serial, time-sharing scheme to reduce workload to a more acceptable level. The decision about which processing strategy to adopt and under what conditions serial or parallel processing occurs will likely depend upon the tasks involved and how those tasks load the operator. Task processing and scheduling strategies are additional complications that any complete time-line analysis model must consider, however, much work remains before they are well understood.

The VACP demand scores were computed each second by summing demands for all active tasks for each of the VACP components; a sliding time-window was not used in this

measure. A cumulative VACP score was computed by simply summing each of the four VACP component values.

The modified W/Index workload model (Sarno and Wickens [14,15]) that was used differed slightly from that originally developed by North and Riley [13]. The modified form of the model that was used is given by the following equation:

$$W_T = \sum_{i=1}^{6}\sum_{t=1}^{M} a_{t,i} + \sum_{i=1}^{6}\sum_{j=i}^{6}\sum_{t=1}^{M-1}\sum_{s=t+1}^{M} c_{ij} \bullet f\left(a_{t,i}, a_{s,j}\right) \qquad (6)$$

where,

W_T is the total workload value;

$a_{t,i}$ is the attentional demand to channel "*i*" due to task "*t*";

$c_{i,j}$ are interference coefficients characterizing the additional load imposed by two tasks competing for common resources;

i,j are indices of the six interface channels: visual input, auditory input, spatial cognition, verbal cognition, vocal output and physical output;

t,s are indices identifying one of "*M*" active tasks;

$f\left(a_{t,i}, a_{s,j}\right)$ is a function that assumes a value of "$a_{t,i}+a_{s,j}$" if both attentional demands are nonzero but it assumes a value of zero if either attentional demand is zero.

This form of the W/Index model was suggested by Sarno and Wickens [14,15] to "...discriminate channel conflict in a task from channel conflict between tasks", a feature of workload they felt was not adequately represented in the original W/Index model of North and Riley. The Within Tasks Demand (the first term of equation 6) represents an estimate of the demands placed on the operator due to all current tasks assuming no conflict between tasks. The Across Tasks Interference (the second term) is an estimate of the additional demands placed on the operator due to interference between tasks. These two components were recorded as well as the resultant modified W/Index score, W_T.

VACP scores and W/Index values were only computed if concurrent, interfering tasks occurred; this resulted in a large number of periods of "zero" workload being predicted as there were many periods of only one or no active discrete tasks. Continuous tasks (such as flying the aircraft) were not included in the Number of Concurrent Tasks count nor were the continuous tasks included in the Time Occupied Parameter.

Analysis Procedure

As both the empirical and analytical assessments of workload were based on subjective evaluations, there was some concern that the data represented only ordinal results (that is, the magnitude of the data values are merely indicators of rank). It is common practice, however, for similar subjective ratings to be treated as at least interval data (that is, the magnitude of the data values quantifies the difference between data points although the magnitude may be relative to some arbitrary reference value as is the Celsius temperature scale). Therefore, both parametric and non-parametric correlation analyses were performed, but most of the assessment is based on the non-parametric analysis results. Pearson Product-Moment correlation (Norman and Streiner [18]) and Spearman ρ rank-order correlation (Siegel and Castellan [19]) were calculated between the empirical workload score and each of the analytical workload measures computed in the task time-line analysis. Similar correlation analyses were performed between each of the empirical TLX-like scores and each of the VACP component values, the cumulative VACP scores, the two W/Index component scores, the cumulative W/Index score and each of the other TLX-like components.

The empirical workload scores were normalized in a similar fashion to that of Navon and Gopher [20] in an attempt to obtain a common measurement interval for all subjects to facilitate the parametric correlation analysis. The scaling relationship used was as follows:

$$s_i = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \tag{7}$$

where,

s_i is the scaled overall workload score;

S_i is the raw overall workload score for a segment of a subject;

S_{min} is the minimum overall workload score over all segments for a subject;

S_{max} is the maximum overall workload score over all segments for a subject.

The workload normalization was done for each subject individually; for subjects that participated in two missions, the normalization procedure was applied to each set of results separately.

The subjective scores for each of the six TLX-style workload components were originally marked on a line with 21 points varying from 0 to 20. The line was doubly-anchored; that is, the endpoints 0 and 20 were accompanied by a descriptive paragraph. The empirical results were scaled to represent a fraction of the overall range of each variable. Initially, the "Own Performance" scale varied from left to right as Best to Worst, whereas the other scales varied from Least to Most in the same direction. The "Own Performance" score was modified prior to the scaling so that a low score indicated a poor performance; this was done for convenience in the analysis so that all of the TLX-like scales had a common format of varying from low (0) to high (20) and has no bearing on the results. The justification for the original performance scale varying from high to low was that it corresponded to a low to high workload as did the other scales; thus, all the scales would combine positively to produce a workload measure. It is doubtful, however, that the subjects actually think about how the scales are combined to form a workload measure and instead consider each scale as an independent entity to be scored.

The analytical study results comprised six sets of data corresponding to the three (Sikorsky, Combined and Individual) task conflict matrices for both the pilots and the observers. Each of these sets of data comprised the nine time-nomalized mission segments, with 3600 data records containing values of the various workload measures for each second of the normalized time-line. The analytical results were reduced somewhat by condensing each of the 3600 workload values to a few descriptive statistics which are described below. While this procedure can reduce the power of correlation analyses, it was not expected to

adversely affect the current analysis as the empirical workload results were represented by single values for each of the perceived TLX-like loads and as well as a single value for the overall workload for a given segment. As it is not known what aspects of the demands placed on a subject are used in formulating the judgement of the perceived loads, amalgamation of the second by second analytical data to obtain values representing a mission segment seemed appropriate.

For each of the Unacceptable Task Conflict, the Unacceptable or Marginal Task Conflict, the Task Conflict Parameter and the Time Occupied Parameter, all segments had minimum values of zero and maximum values of 100%, making the maximum and minimum values of these descriptive measures of little diagnostic value for comparison with the workload scores. Thus, the data were reduced to the following measures over each segment: the average of each measure; average of all non-zero values; the frequency of occurrence of a parameter attaining a value greater than zero but less than or equal to 25% (referred to as "1st quartile"); the frequency of occurrence of a value greater than 25% but less than or equal to 50% (2nd quartile); the frequency of occurrence of a value greater than 50% but less than or equal to 75% (3rd quartile); the frequency of occurrence of a value greater than 75% but less than or equal to 100% (4th quartile). While the sum of each of the products of frequency of occurrence times the probability of overload reproduces the mean probability of an overload condition, it was hoped that the additional detail would provide some further insight into how subjects perceive workload since a uniform 25% probability of overload may result in a different perception than would a 100% probability of overload for 25% of the time.

The Number of Concurrent Tasks data were reduced to the following measures: the maximum number of discrete, concurrent tasks; the mean number of concurrent tasks; the number of times a single task occurred in a segment; the number of times two concurrent tasks occurred; the number of times three concurrent tasks occurred; the number of times four concurrent tasks occurred; and, the number of times five or more concurrent tasks occurred.

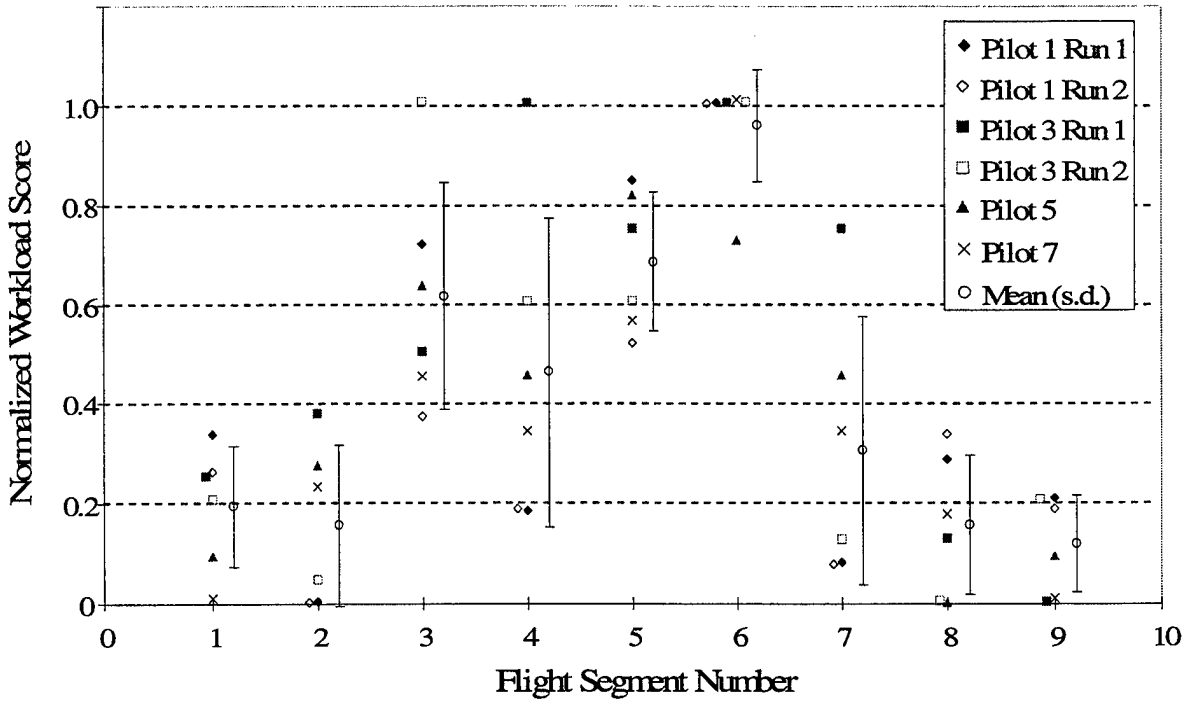
The following were computed for the VACP and W/Index models (both for the individual components of each model and for the cumulative values): the maximum value of the measure attained in a segment; the mean value over a segment; the mean of all non-zero values over a segment; the fraction of time that the measure had a zero value in a segment.

Results

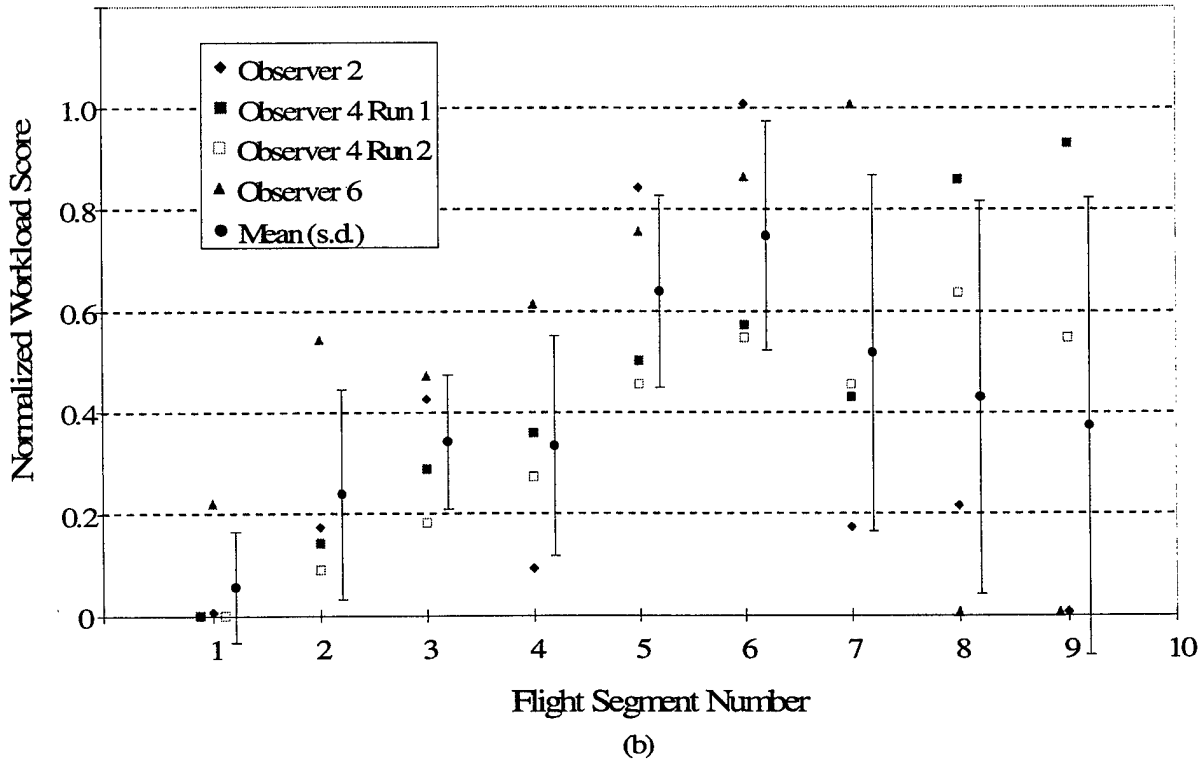
The subjects' normalized workload values for each flight segment are shown in Figure 1. The individual datum points as well as the segment mean and standard deviation are included in the figure, reflecting the concern that the data may be only ordinal. There is considerable scatter in the perceived workload within each segment; both the range and standard deviation varied between 20% and 90% of the total extent of the subjective load experienced during the exercise.

The data showed slightly different trends for the pilots than for the observers. The range in any segment was somewhat smaller for the pilots than it was for the observers. Pilots reported the greatest workloads were experienced in segments 3 through 6 with segments 5 and 6 being the more consistently higher scoring; greater observer workloads were experienced in segments 5 through 9, although again, segments 5 and 6 seemed to be more highly rated than segments 7 through 9.

Rank order correlation (Spearman's ρ) between the individual pilot's scores varied from a low of 0.3 to a high of 0.8 but typical values were about 0.65, with most correlations being statistically significant ($\alpha<0.02$). These results are similar to other correlations between analytical and empirical workload values reported in the literature for complex scenarios (Riley et al. [21], Iavecchia et al. [22], Bateman and Thompson [23]). The individual observer's workload scores did not correlate as well as the pilots', with Spearman ρ values varying from -0.5 to 0.3 with much lower statistical significance ($\alpha\sim0.2$). Three subjects participated in two missions: two pilots and one observer. Pilot 1 and Observer 4 both had highly correlated workload scores between their two missions ($\rho>0.9$, $\alpha=0.0001$) while the workload scores from the missions of Pilot 2 had only a moderate correlation ($\rho=0.6$, $\alpha=0.03$).



(a)



(b)

Figure 1. Normalized subjective workload values versus flight segment number for all subjects as well as the group mean (standard deviation) in each flight segment for (a) pilots and (b) observers.

Correlation coefficients between the empirical normalized workload scores and the various analytical measures are presented in detail in Appendix A with a sample shown below:

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Across Task Interference (mean of nonzero values)	**0.70**	**0.71**	0.69	**0.70**	0.69	**0.71**
Unacceptable Task Conflict (3rd quartile)	**0.70**	**0.77**	0.53	**0.70**	0.53	0.69
Unacceptable or Marginal Task Conflict (3rd quartile)	0.69	0.47	0.54	0.41	0.49	0.39

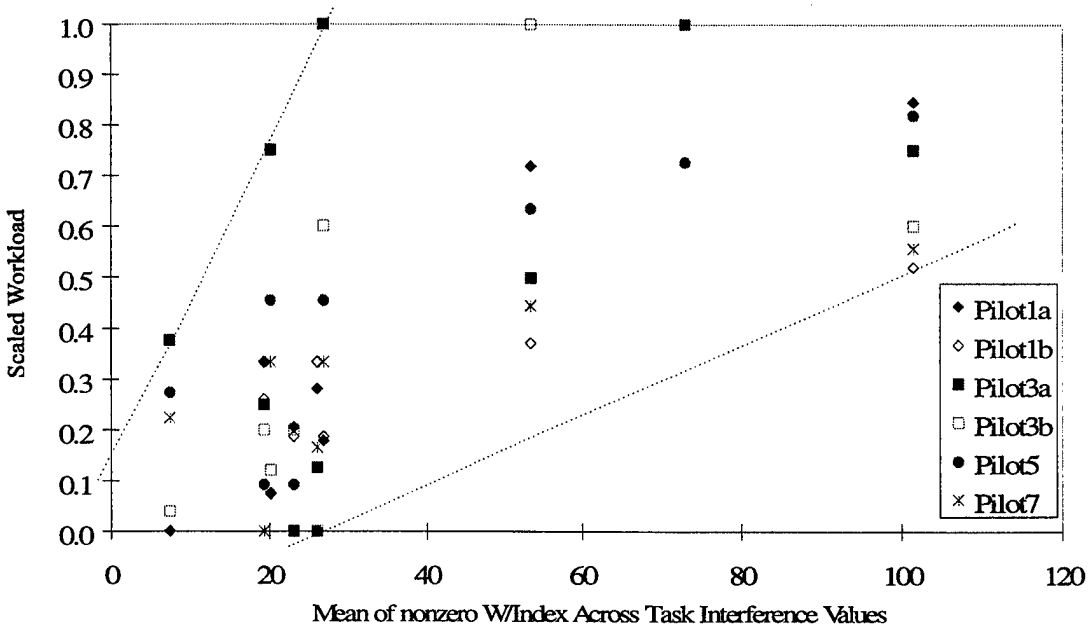
Both non-parametric (Spearman's rank-order ρ-value) and parametric (Pearson's r-value) values are presented for each of the three sets of results arising from the use of the different conflict matrices. Results that had a correlation greater than 0.7 are shown in bold and statistical significance at the 0.01 level was obtained for all values greater than approximately 0.4. None of the correlation coefficients is outstanding and only a few of the analytical measures explained greater than 50% of the variance in the empirical workload scores. The correlations between the pilots' empirical workload scores and the analytical workload values were generally greater than that found for the observers

The results of each of the six individual TLX-like measures were compared to the analytical measures and these results are presented in Appendix B with a sample shown below for two of the measures:

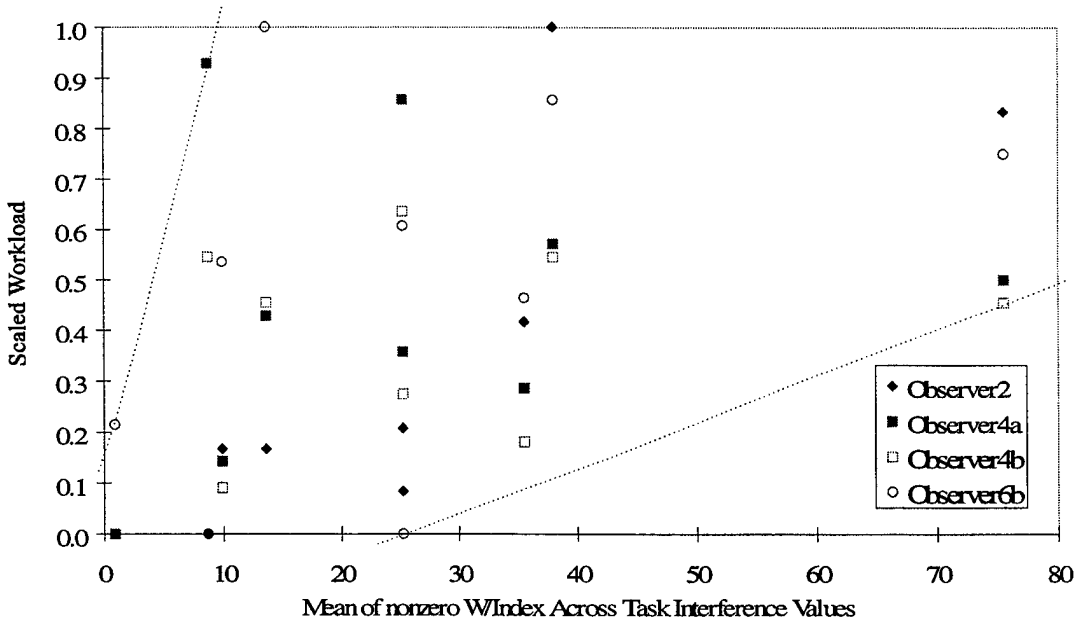
Pilots: Individual Conflict Matrix					
Activity Level			**Anxiety Level**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Auditory Attent. Demand mean	0.73	<.0001	Cognitive Attent. Demand max	0.69	<.0001
Cognitive Attent. Demand max	0.71	<.0001	Across Task Interference mean nonzero	0.67	<.0001
Cumulative VACP mean nonzero	0.70	<.0001	Auditory Attent. Demand mean	0.65	<.0001

The comparison is grouped by subject group (pilot or observer) and according to the set of conflict matrices used. The results are further grouped under each of the six empirical measures showing the Spearman rho correlation coefficient and the associated significance level, α, for each of the comparisons.

Figure 2 shows an example of the correlation between the empirical data for the pilots' workload and that predicted using one of the better-correlating, multi-dimensional analytical measures, the mean of all nonzero values of the W/Index Across Task Interference.



(a)



(b)

Figure 2. Plot of the scaled workload scores for (a) pilots and (b) observers versus the mean of all nonzero values of the W/Index Across Task Interference values. The dashed lines indicate the extent of the spread of the data.

Discussion

Some of the scatter in the empirical workload values shown in Figure 1 can likely be attributed to a number of uncontrolled variables such as: external stressors that varied between missions (weather, field exercise details, etc.); differences between the subjects' capabilities; differences in the same subject between missions. The accuracy of the subjective measurement technique may also be a significant factor in the recorded workload variance. Taking factors such as these into account when attempting to design a field experiment or to normalize the empirical workload data from field experiments would be a formidable task; indeed, it may not be possible to accomplish such a transformation at all in a rigorous fashion for field exercises in which collection of workload data is secondary to a primary objective such as training, control of a vehicle or safety.

In most segments, the range of responses was smaller for the pilot group than for the observer group. This difference could indicate a more uniform background or experience level for the pilot group, however, this is speculation as details on the subjects' experience level were unavailable. Alternatively, the pilots tasks are more rigorously constrained by the flight environment than are the observers tasks which could lead to a more uniform demand placed on the pilots than on the observers (Mack [24]). If the observers do have greater freedom to perform their tasks, individual abilities and workload strategies may be more apparent than in the pilots' results. Some of the difference, perhaps even most of the difference, may result from the subjective measurement procedure and the observed difference may simply be a typical variance, exacerbated by the small number of subjects in each group.

The low to moderate correlation between the empirical group workload scores in Appendix A is likely attributable in large measure to individual differences between subjects as well as variations between missions as previously mentioned. Discussion of these results is broken-down into uni-dimensional and multi-dimensional workload measures as well as

the contribution of the components of the multi-dimensional workload (Appendix B), both analytical and empirical measures.

Uni-dimensional Workload Measures

Of the single-dimension workload measures, Unacceptable Task Conflict seemed to have the most consistent and greatest correlation with workload across all three sets of conflict matrices for both pilots and observers, although, at best, this measure only captured 50% of the pilot workload variance and about 25% of the observer workload variance. This correlation was evident with the frequency of occurrence of either 25-50% or 50-75% probability of an Unacceptable Task Conflict: that is, the incidence of intermediate or moderately high probability of an Unacceptable Task Conflict seems to be indicative of the demands experienced by the subjects. Unfortunately, these two categories cover such a wide range of probability of overload as to make their use in prediction of limited value. Review of the data indicates that, overall, the incidence of each of the 25-50%, 50-75% and the 75-100% probability of Unacceptable Task Conflict were not markedly different in any given segment although the frequency of occurrence in each category did vary between segments; segment 5 seemed to have the highest incidence of Unacceptable Task Conflict.

Surprisingly, a weaker correlation was found between the workload and the incidence of a 75-100% probability of an Unacceptable Task Conflict which begs the question: are these descriptive statistics meaningful for this parameter? It is possible that this result may be a reflection of the limitation of the network model used; if such a high load occurred in practice, the subject may modify their processing strategy, say accepting a greater error for less effort, resulting in a lower perception of workload than predicted from the network model. Continuous, instantaneous empirical results are unavailable (due to the nature of the data) for comparison with the instantaneous analytical values necessitating the reduction of the analytical results to something that might compare with the empirical results; unfortunately, not all of the descriptive statistics that can be derived are meaningful or useful, and it may be that the incidence of occurrence of the various Task Conflict measures may be

one of those that are inappropriate. No satisfactory conclusion was reached regarding the validity of these measures.

The mean value of the Unacceptable Task Count showed no substantial correlation with workload for the pilots and captured less than 20% of the workload variance for the observers. The combined Unacceptable and Marginal Task Conflict parameter showed similar results to the Unacceptable Task Conflict parameter although with slightly poorer correlation; it would appear that including the occurrence of Marginal Task Conflicts with the Unacceptable Task Conflict measure of workload results in a poorer predictive capability. The Task Conflict Parameter did not correlate well with workload for either the pilots or the observers and at best captured only 25% of the workload variance.

The incidence of the Number of Concurrent Tasks greater than four was also indicative of high workload for pilots, although less so for observers; the incidence of Number of Concurrent Tasks greater than four was approximately four times as great for the pilots as it was for the observers. It should be noted, however, that for the pilot group, the incidence of the two concurrent tasks (Number of Concurrent Tasks 2) was correlated to workload nearly as well as more than four concurrent tasks although it was a negative correlation. This negative correlation is likely something of an artefact. It is likely that the execution of two concurrent tasks was often routinely handled without excessive workload but as the incidence of more than two concurrent tasks increased (and hence the incidence of only 2 tasks decreased) with a concomitant increase in workload, a negative correlation of workload and incidence of 2 concurrent tasks might occur. Neither of the above two measures was found to capture more than about 40% of the workload variance. The mean and maximum Number of Concurrent Tasks in a segment only managed to capture between 20% and 25% of the workload variance respectively. The incidence of one, three or four Concurrent Tasks did not correlate well with workload, with any one of these measures capturing less than about 25% and 15% of the workload variance for the pilot and observer respectively.

The incidence of the Time Occupied Parameter having between 75% and 100% probability captured approximately 30% of the workload variance for pilots and about 20% of the observer workload variance. None of the other Time Occupied Parameter measures managed to capture a substantial amount of the workload variance for either the pilots or the operators.

Multi-dimensional Workload Models

Of the two multi-dimensional analytical workload measures, the W/Index score seemed to correlate best with the empirical workload scores. The maximum value of the W/Index parameter had the greatest correlation with the pilot workload scores ($\rho \approx 0.65$) although the mean of the non-zero values of the W/Index score ranked higher more often across both subject groups. The mean of the non-zero values of the W/Index score had a similar correlation for the pilots as the Maximum W/Index score but showed a better correlation for the observers ($\rho \approx 0.43$ versus $\rho \approx 0.30$ for the maximum W/Index value). None of the multi-dimensional analytical workload measures succeeded in predicting more than about 40% of the empirical workload variance; whether this is due to limitations of the modelling, the small number of subjects or the empirical metric is unclear, although it is likely that each of these factors were at fault to some degree. For the pilot group, there was little difference between mean and maximum W/Index values of the predictive workload correlation with empirical workload over a segment while for the observer group, the mean values seemed to capture somewhat more of the perceived-workload variance than did the maximum W/Index value.

The mean value of the Across Task Interference component of the W/Index calculation (shown in Figure 2) was found to have about the best correspondence with the perceived workload although several of the VACP components, notably the mean Auditory Attentional Demand and the mean, non-zero value of the Cognitive Attentional Demand were comparable in

performance. The correlation between these measures and the empirical workload score was approximately 0.7 or approximately 50% of the workload variance. Figure 2 shows the extent of the scatter, however, the graphs may be somewhat deceiving as the range of the scaled workload does not span the entire range possible for workload; one might expect both greater and lesser levels of workload than those encountered during the experiment. That the whole was less than some of the individual parts of the analytical workload models suggests that a simple summing of components of the multidimensional workload measures may not be the most effective approach, but that a nonlinear weighting of component contributions or conditional relationships similar to Fuzzy Set Inference approaches may be more suitable for predicting perceived workload.

The incidence of zero values for either the VACP or the W/Index model did not correlate well with workload. It was originally thought that that there might be a significant negative correlation of the incidence of low activity with workload. The lack of correlation may be an artefact of a task-based execution of the task network, since in the analysis, when one task finished another was started; that is, there is no idle time. Zero values were, however, the predominant value for all of the analytical measures since there were periods of little activity and calculation of demands were made only when two or more conflicting tasks occurred.

The maximum value of the Cumulative VACP measure, in most cases, correlated almost as well as the maximum W/Index score, both for pilot and observer workload predictions. The best correlation was with the Individual Conflict Matrices for both groups, where the correlation captured approximately 36% of the workload variance. The mean of the non-zero VACP scores did not correlate quite as well as the mean non-zero W/Index score except for the Individual Pilot matrices in which this measure captured approximately 46% of the pilot workload variance; the similar measure for the observer group using the Individual Observer conflict matrices captured 16% of the observer workload variance. Other measures of the VACP results showed little correlation with the empirical workload data.

Multi-dimensional workload components

The individual empirical TLX-like scores were most often found to have the greatest correlation with other TLX-like scores (Table 1) except for the subjects' assessment of their own performance (Performance) which did not substantially correlate with any of the empirical or analytical measures. The high inter-measure correlation of the TLX-like scores could result either from an inherent correlation between measures or because of the characteristics of the range of tasks selected for assessment. If, in the later case, the high degree of inter-measure correlation is a reflection of the demands of the selected tasks, then the inter-correlations are appropriate. If the inter-measure correlation is due to intrinsic correlation of the measurement categories, then there will be redundancy in the measurement method which is desirable, to a degree, lending confidence to the measured value; the down-side to such redundancy is that some other, more independent measure may be ignored due to practical limitations in collecting the data, restricting the generality of the assessment procedure. In the development of the TLX rating scales, Hart and Staveland (1988) also note substantial correlation between several of the preliminary scales considered and used this information to eliminate some scales in the final selection of scales. Even then, however, the final scales showed moderate correlations between some of the scales (for example, $r \approx 0.76$ between Effort and Mental Demand) somewhat lower than the inter-scale correlations found in this study. Since complete separation of the scales is unlikely or at least would be a difficult undertaking, Hart and Staveland seemed content to rely on the practical position that for significantly-correlated factors, "...shared contribution to a weighted estimate of overall workload is simply enhanced, reflecting the actual situation." Without further analysis specifically examining the underlying correlation within the TLX or TLX-like component categories, it is difficult to say whether this position is justified or if the overall measurement scheme is inefficient and could be improved by a selection of different categories.

Table 1. Inter-correlation of the empirical TLX-like scores for the pilots and the observers. The upper-triangular matrix represent Spearman rank order ρ-values while the lower triangular matrix represent Pearson product-moment r-values.

Pilots	Activity Level	Anxiety	Mental Effort	Performance	Physical Effort	Task Complexity
Activity Level		0.87	0.84	0.05	0.64	0.92
Anxiety	0.93		0.86	0.12	0.60	0.88
Mental Effort	0.89	0.90		0.11	0.63	0.87
Performance	0.03	0.15	0.08		0.21	0.13
Physical Effort	0.75	0.73	0.73	0.19		0.62
Task Complexity	0.93	0.91	0.90	0.16	0.75	

Observers	Activity Level	Anxiety	Mental Effort	Performance	Physical Effort	Task Complexity
Activity Level		0.70	0.66	-0.30	0.18	0.92
Anxiety	0.72		0.34	-0.13	0.05	0.72
Mental Effort	0.66	0.33		-0.16	0.18	0.60
Performance	-0.25	-0.16	-0.17		-0.17	-0.23
Physical Effort	0.21	0.01	0.24	-0.19		0.06
Task Complexity	0.92	0.74	0.60	-0.22	0.08	

Tables of the correlation of empirical TLX-like scores with the analytical measures are presented in Appendix B. The pilot group data is given first, showing the results of each of the three sets of conflict matrices, followed by similar tables for the observer group data. The correlation results are Spearman-ρ rank-order correlation; a two-tailed t-test was used to test the significance of the correlation (Siegel and Castellan [19]).

The pilots' Activity Level score had its strongest correlation, across all three conflict matrices, with the mean Cognitive Attentional Demand (ρ~0.7). The mean Auditory Attentional Demand had a similar correlation for the Combined and Individual Conflict matrices, however, the correlation was weaker for the Sikorsky Conflict matrices (ρ~0.4). Somewhat counter-intuitively, the pilot Activity Level had a negative, although weak,

correlation with Physical Attentional Demands, suggesting that the pilots associated the concept of Activity Level more with aspects such as decision making and situation awareness rather than the mechanics of flying the mission which may have become a "Skill" in Rasmussen's [25] taxonomy; this association may change, however, if they are flying in adverse weather requiring higher levels of physical attention. The mean of the non-zero interference terms of the W/Index relationship (Across Task Interference mean nonzero) also had a fairly high correlation with Activity Level ($0.6 \leq \rho \leq 0.7$) suggesting that conflicts in competition for resource channels contributes a significant amount to the perception of Activity, however, the maximum value of the Within Task Demand of the W/Index relationship had a correlation that was only slightly weaker ($0.5 \leq \rho \leq 0.6$). Combining the analytical components into either the total W/Index or cumulative VACP measures did not result in a better correlation with the Activity Level measure.

The pilots' Anxiety Level score did not seem to correlate consistently better with any single measure across all three of the conflict matrices. The Anxiety Level did have a moderate correlation ($0.55 \leq \rho \leq 0.67$) with the mean value of the Across Task Interference terms of the W/Index relationship and similar correlation with the Cognitive Attentional Demand. There was a comparable, but negative, correlation with the incidence of zero scores in the Physical Attentional Demand ($-0.42 \leq \rho \leq -0.54$). It is possible that Anxiety is associated with periods of high mental activity which, for the tasks encountered, did not have a concurrent high physical activity level or it may suggest that the pilots did not feel quite as stressed while they were physically busy. This observation should be tempered with the further observation that there was a comparable negative correlation between the incidence of zero values of the Auditory Attentional Demand ($0.4 \leq \rho \leq 0.5$) which could simply mean that when there was relatively little going on, the pilots did not feel stressed, which seems rather intuitive if there is no anticipation of impending high-demand activities.

The pilots' Mental Effort scores had their greatest correlation with the mean of the non-zero values of the Across Task Interference term of the W/Index relationship ($\rho = 0.74$) with the Combined and Individual Conflict matrices; with the Sikorsky Conflict matrix,

correlations were substantially lower between all measures and Mental Effort ($\rho \leq 0.43$). The mean values of the mean Auditory and maximum Cognitive Attentional Demands had comparable correlations ($\rho \sim 0.65$) with Mental Effort to the mean non-zero Across Task Interference terms which may suggest that more of the Mental Effort was associated with receiving auditory messages, processing those auditory messages and the conflict that arises due to other concurrent tasks interfering with the auditory demands. A similar observation could be made using the mean non-zero value of the Visual Attentional Demand, although the correlation is slightly lower ($\rho \sim 0.56$).

As previously noted, the pilots' assessment of their own Performance did not correlate well with any of the analytical measures ($\rho \leq 0.26$). The lack of correlation might be explained if Performance was associated with relative personal achievement rather than an absolute measure of how well the task was completed. The assessment of Performance might well be an indicator of the subject's cognitive and physical state, and thus this measure might be better employed to modify some of the other empirical measures in order to compare with a normative analytical model's results; the same approach might also be taken with the Anxiety Level measure.

The pilots' Physical Effort scores were only moderately well correlated with the analytical measures and the correlation was stronger using the Combined or Individual Conflict matrices rather than the Sikorsky matrices. The correlation seems to be strongest with the W/Index model and its two components ($\rho \sim 0.55$), however, the cumulative VACP score along with the Visual and Auditory Attentional Demands each had comparable correlation coefficients to that for the W/Index model.

The pilots' Task Complexity scores correlated best with the mean nonzero Across Task Interference measure ($\rho \sim 0.7$) for both the Individual and the Combined Conflict matrices but had a similar correlation with the mean nonzero Within Task Demand for the Sikorsky Conflict matrix although the correlation with the Sikorsky Conflict matrix mean nonzero Across Task Interference was comparable ($\rho \sim 0.6$). The maximum value of the

Cognitive Attentional Demand had a similar correlation ($\rho \sim 0.6$) for the Individual and Combined Conflict matrices but less so for the Sikorsky Conflict matrix ($\rho \sim 0.5$); instead, the mean nonzero values of the Cumulative VACP and Cognitive Attentional Demand parameters correlated better with Task Complexity ($\rho \sim 0.7$) for the Sikorsky Conflict matrix.

The observers' Activity Level scores correlated best with the mean nonzero values of the Cumulative VACP and the Cognitive Attentional Demand ($\rho \sim 0.7$) across all three sets of conflict matrices, similar to that found for the pilot group. The mean nonzero Within Task Demand correlation was also high ranking with a correlation coefficient of $\rho \sim 0.7$ for the Combined and Sikorsky Conflict matrices, although the Individual Conflict matrix yielded a lower correlation for this value ($\rho \sim 0.6$); the mean nonzero value of the Across Task Interference term was correlated somewhat weaker ($0.5 \leq \rho \leq 0.6$). This suggests that the observer group considered Activity to be associated more with cognitive activities involving the load imposed by the individual tasks rather than competition for cognitive resources between tasks.

The observers' Anxiety level scores correlated most strongly with the same measures as did the Activity level, although the correlation was about 10 to 15% weaker for the higher correlations. The mean non-zero value of the Within Task Demand of the W/Index relationship, the mean non-zero value of the Cumulative VACP relationship and the mean non-zero value of the Cognitive Attentional Demand were higher correlating factors with Anxiety Level ($\rho \sim 0.6$) suggesting that Anxiety was predominantly associated with the load imposed by mental tasks.

The observers' Mental Effort scores correlated poorly with all measures; the strongest correlation was with the mean nonzero values of the Cognitive Attentional Demand and the Cumulative VACP scores ($\rho \sim 0.4$); the mean nonzero value of the Within Task Demand had a similar with the Combined and Sikorsky Conflict matrices but less so with the Individual Conflict matrix ($\rho \sim 0.3$). This suggests that the analytical methods are not adequately

representing the demands made on the cognitive effort for the observer group or the models are predicting some variable different from what the Mental Effort score is thought to represent.

The observers' assessment of their own Performance, as in the case with the pilot group, did not correlate with any of the analytical multidimensional workload measures or their components.

Although there was a somewhat stronger correlation of the analytical results with the observers' Physical Effort scores than with their Performance scores, the correlation was very weak ($\rho \leq 0.4$). Correlation with the various Physical Attentional Demand values varied considerably between the three sets of conflict matrices ($0.1 \leq \rho \leq 0.4$) and only in the Individual Conflict matrix results was the maximum value of the Physical Attentional Demand measure close to being the best correlate with the Physical Effort score. Generally, the mean nonzero values of the Cumulative VACP and the Cognitive Attentional Demand correlated best with the Physical Effort score ($\rho \sim 0.4$).

The observers' Task Complexity score correlated most strongly with the mean nonzero value of the Cognitive Attentional Demand ($\rho \sim 0.7$) and a similar correlation was found with the Cumulative VACP value. The mean nonzero value of the Within Task Demand also had a stronger correlation with the Sikorsky and Combined Conflict matrices ($\rho \sim 0.7$) but was slightly lower using the Individual Conflict matrix ($\rho \sim 0.6$); the mean nonzero value of the Across Task Interference correlation with the Task Complexity was comparable to but slightly less than the Within Task Demand ($0.5 \leq \rho \leq 0.6$). This suggests that Task Complexity is most associated with the individual cognitive demands made by the tasks but that interference due to competition between the tasks for the cognitive resources is a close second.

Conclusion

The conclusions that can be drawn from the analysis described in this report should be viewed in the light of the small subject sample size of the field exercise. The small number of subjects available for the study significantly limits generalization of the conclusions and perhaps even compromizes the validity of the conclusions.

Of the overall workload measures, the uni-dimensional measures were found to capture the greatest portion of the empirical workload variance although these measures provide little detail in what is actually causing the overload and at best only capture 50% of the workload variance. While the multi-dimensional workload measures may provide greater detail about what is causing high workloads, they do not seem to be capturing a great deal of the workload variance to begin with.

Similarly, the TLX-like workload component scores did not correlate very well with any of the analytical workload measures. At best, half of the variance in the empirical measures was captured by any of the analytical measures. Surprisingly, the empirical Activity Level had a very low correlation with the analytical VACP measure of Physical Attentional Demand for the pilot group, perhaps suggesting that the pilots associated the concept of Activity more with things such as decision making and situation awareness rather than the mechanics of flying the aircraft.

The analytical TLX-like workload component measures had a significant degree of inter-correlation, much more so than the correlation with any of the analytical measures. It is unclear whether this is due to the tasks involved in the analysis or if these measures are inherently correlated. If the measures are intrinsically correlated for all tasks likely to be encountered in an analysis, then a more efficient measurement approach might be adopted by eliminating some of the highly inter-correlated variables and perhaps substituting some more independent measures considered to be of lesser importance. Such an approach may capture

more of the overall workload variance although it may compromize some of the diagnostic capability of the measurement scheme.

Although the correlations are low and only half the workload variance was captured, the models may still be useful; it has been said that "a model need not be right to be useful". What it does mean, however, is that the models should not be treated as "black-boxes" and applied indiscriminately. In order for the workload models and the task network analyses to produce reliable and useful information, they must be applied by a knowledgeable analyst having a good understanding of the models limitations and the models should be validated whenever they are applied in a novel scenario. The models are not at a level of sufficient maturity that they can be released as a "textbook tool" for use by the uninitiated. Needless to say, more research is required on the fundamental causes of workload as well as means of measuring workload in a psychometrically sound approach.

The results of the investigation do suggest a few points that should be considered for developing experiments that might be used for validating task network workload models. It was found that the analysis of the results by parametric or non-parametric techniques did not produce significantly different results; this may be peculiar to the type of investigation considered here, so if there are any concerns that some of the data may be simply ordinal, then non-parametric analysis techniques are recommended. Since the empirical analysis is so time-consuming, it is recommended that mission segments should be selected such that each is relatively short and the level of workload remain approximately constant throughout the duration of the segment. Although statistical significance was achieved for many of the correlations in this investigation, the sample size should really be larger to increase the generalization of, and the confidence in, the conclusions. In the modelling, it would be instructive to include the continuous tasks in the VACP and W/Index measures of workload; while omission of continuous tasks from Time Occupied measures is understandable, such a step is not required from more complex measures. It would also be desirable to include workload values when even only one task is present which might result in improved correlation at the lower end of the workload range. One final issue dealing with the

analytical approach has to do with the presentation of data to the analyst; the data should be saved in a less processed state, leaving out much of the on-line averaging so that the analyst can, if desired, generate descriptive statistics after the fact.

In summary, the analytical methods were only moderately successful in predicting the empirical measure of workload, whether at the overall level or at the component level. Both the analytical and empirical approaches seem to have significant limitations and would be well served by additional research into their fundamentals. In spite of the limitations, the models may still be useful in the evaluation of new equipment designs or procedures. Again, these conclusions should be considered with full knowledge that they are based on what is likely an inadequate subject sample size.

References

1. Hendy, K.C., Hamilton, K.M., Landry, L.N., (1993) **Measuring subjective workload: when is one scale better than many?** *Human Factors*, 35(4), 579-601.

2. Shaffer, M.T., Hendy, K.C. and White, L.R., (1988) **An empirically validated task analysis (EVTA) of low level army helicopter operations.** *Proceedings of the Human Factors Society - 32nd Annual Meeting, Santa Monica, CA.,* 178-182.

3. Hart, S.G., Battiste,V., Chesney, M.A., Ward, M.M. and McElroy, M., (1986) **Comparison of workload, performance, and cardiovascular measures: Type A personalities vs Type B.** Working Paper. Moffett Field, CA: NASA Ames Research Center.

4. Haworth, L.A., Bivens, C.C. and Shively, R.J., (1986) **An investigation of single-piloted advanced cockpit and control configuration for nap-of-the-earth helicopter combat missions tasks.** *Proceedings of the 1986 Meeting of the American Helicopter Society*, Washington, D.C. 657-672.

5. Gopher, D. and Braune, R., (1984). **On the psychophysics of workload: Why bother with subjective measures?** *Human Factors*, 26(5), 519-532.

6. Hart, S.G. and Staveland, L.E., (1988) **Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.** In P.A. Hancock., N. Meshkati, (Eds.) *Human Mental Workload*, pp139-183 Elsevier Science Publishers, B.V. (North Holland).

7. Youngson, G.A., McKay, D. and Kobierski, R.D., (1995) **VACP validation programme final report.** Canadian Marconi Company, Document Number 1000-1059, DSS Contract No. W7711-2-7180/01-XSE, Defence and Civil Institute of Environmental Medicine Contract Report.

8. Anonymous, (1993) **Systems Operator Loading Evaluation Facility (SOLE), Version 3.0, Methodology Documentation**. Canadian Marconi Company, Avionics Division, Systems Engineering Group, Defence and Civil Institute of Environmental Medicine Contract Report.

9. Seifert, D.J., Hann, R.L. and Chubb, P.G., (1978) **Simulation using SAINT: A user-oriented instruction manual.** Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, Report AMRL-TR-77-61.

10. Wortman, D.B., Duket, S.D., Seifert, D.J., Hann, R.L. and Chubb, G.P., (1978) **Simulation using SAINT: a user-oriented instruction manual.** Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, AMRL-TR-77-61, July.

11. McCraken, J.H. and Aldrich, T.B. (1984). **Analyses of selected LHX mission functions: Implications for operator workload and system automation goals.** US Army Research Institute Aircrew Performance and Training, Technical Report No. MDA903-81-C-0504 ASI479-024-84

12. Schuck, M.M., (1996) **Development of equal-interval task rating scales and task conflict matrices as predictors of attentional demand.** *Ergonomics*, 39(3), 345-357.

13. North, R. A. and Riley, V.A. (1989). **W/INDEX: A predictive model of operator workload.** In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, L. van Breda (Eds.) *Applications of human performance models to system design.* pp. 81-90. New York, Plenum Press.

14. Sarno, K.J. and Wickens, C.D., (1991) **The role of multiple resources in predicting time-sharing efficiency: an evaluation of three workload models in a multiple task setting**. Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-

Champaign, ARL-91-3/NASA A^3I-91-1, NASA Ames Research Center contract number NASA NCC 2-632.

15. Sarno, K.J. and Wickens, C.D., (1995) **Role of multiple resources in predicting time-sharing efficiency: evaluation of three workload models in a multiple-task setting.** *International Journal of Aviation Psychology*, 5(1), 107-130.

16. Hamilton, B. (1984) **ARTI Task I Workload Methodology**. United Technologies, Sikorsky Aircraft, Document Number SER-760502, DAAK51-84-C-0007.

17. Liao, J. and Moray, N. (1993) **A simulation study of human performance deterioration and mental workload.** *Le Travail humain*, 56(4), 321-344.

18. Norman, G.R. and Streiner, D.L. (1994) **Biostatistics: the bare essentials.** Mosby, St.Louis.

19. Siegel, S. and Castellan, N.J., jr. (1988) **Nonparametric statistics for the behavioral sciences.** 2nd Ed., McGraw-Hill Book Company, New York.

20. Navon, D. and Gopher, D. (1979) **On the economy of the human-processing system.** *Psychological Review*, 86(3), 214-255.

21. Riley, V., Lyall, E., Cooper, B. and Wiener, E. (1993) **Analytic methods for flight-deck automation design and evaluation, Phase One Report: Flight crew workload prediction.** FAA Contract Number DTFA01-910C-00039, Department of Transportation, Federal Aviation Administration, Flightcrew Systems Research Branch, ARD 210, Washington, D.C., Honeywell Technology Center, Minneapolis, MN.

22. Iavecchia, H.P., Linton, P.M., Bittner, A.C. and Byers, J.C. (1989) **Operator workload in the UH-60A Black Hawk: Crew results vs. TAWL model predictions**. *Proceedings of the Human Factors Society 33rd Annual Meeting, Denver CO.*, 1481-1485.

23. Bateman, R.P. and Thompson, M.W. (1986) **Correlation of predictied workload with actual workload measured using the Subjective Workload Assessment Technique.** SAE Aerotech '86.

24. Mack, I. (1997) Personal communication.

25. Rasmussen, J. (1986) **Information processing and human machine interaction**, North Holland Publishers, New York.

Appendix A.

Correlation of empirical overall workload scores with analytical workload measures.

The following tables present the correlation of the empirical, scaled workload scores with the predicted, analytical workload measures. "Sikorsky", "Combined" and "Individual" refer to the task conflict matrices used in the analysis. The data for the pilot group is presented first, followed by the data for the observer group. Spearman's rank-order, non-parametric correlation coefficients are listed for each of the sets of matrices under the column headings titled "ρ-values" while Pearson's product-moment correlation coefficients are listed under the column headings "r-value". Correlation coefficients greater than or equal to 0.7 (corresponding to a variance of at least approximately 50% with overall workload) are shown in bold type. For the pilot group, correlation values greater than |0.35| are statistically significant at the 0.01 level; correlation values greater than |0.27| are statistically significant at the 0.05 level. For the observer group, correlation values greater than |0.42| are statistically significant at the 0.01 level; correlation values greater than |0.33| are statistically significant at the 0.05 level. Significance testing is based on a 2-tailed t-test with 52 degrees of freedom for the pilot group and 34 degrees of freedom for the observer group.

The second-by-second analytical values of the "Unacceptable Task Count", "Unacceptable plus Marginal Task Count", "Task Conflict Parameter" and "Time Occupied Parameter" in each segment were reduced for analysis as follows: the mean of non-zero values; 1st quartile representing the incidence of non-zero probabilities of overload up to 25%; 2nd quartile representing the incidence of 25 to 50% probability of overload; 3rd quartile representing the incidence of 50 to 75% probability of overload; 4th quartile representing the incidence of 75 to 100% probability of overload.

Each of the component values as well as the overall values of the W/Index and VACP models were reduced to the following measures for the correlation analysis: maximum value in a segment; the mean value during a segment; the mean of non-zero values during a segment; the incidence of zero values during a segment.

Pilot data

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Across Task Interference (mean of nonzero values)	**0.70**	**0.71**	0.69	**0.70**	0.69	**0.71**
Unacceptable Task Conflict (3rd quartile)	**0.70**	**0.77**	0.53	**0.70**	0.53	0.69
Unacceptable or Marginal Task Conflict (3rd quartile)	0.69	0.47	0.54	0.41	0.49	0.39
Unacceptable Task Conflict (2nd quartile)	0.68	0.37	0.69	0.42	0.66	0.45
Auditory Attentional Demand (mean)	0.68	**0.71**	0.68	**0.71**	0.68	**0.71**
Number of Concurrent Tasks (n>4)	0.66	0.63	0.66	0.63	0.66	0.63
Cognitive Attentional Demand (mean nonzero)	0.65	**0.70**	0.50	0.64	0.50	0.63
W/Index (maximum value)	0.59	0.69	0.63	0.60	0.58	0.59
Auditory Attentional Demand (maximum value)	0.58	0.65	0.58	0.65	0.58	0.65
Cognitive Attentional Demand (maximum value)	0.57	0.67	0.66	**0.70**	**0.72**	**0.70**
W/Index (mean of nonzero values)	0.57	0.69	0.58	0.68	0.58	**0.70**
Across Task Interference (mean)	0.57	0.55	0.54	0.54	0.44	0.54
Number of Concurrent Tasks (mean)	0.56	0.68	0.56	0.68	0.56	0.68
Time Occupied Parameter (4th quartile)	0.55	0.63	0.55	0.63	0.55	0.63
Across Task Interference (maximum value)	0.54	**0.71**	0.65	**0.70**	0.62	0.65
Within Task Demand (mean of nonzero values)	0.53	0.67	0.54	0.68	0.59	**0.73**
Number of Concurrent Tasks (maximum value)	0.52	0.62	0.52	0.62	0.52	0.62
Task Conflict Parameter (2nd quartile)	0.51	0.50	0.46	0.55	0.42	0.56
Unacceptable Task Conflict (4th quratile)	0.51	0.66	0.41	0.60	0.39	0.58
Visual Attentional Demand (mean of nonzero values)	0.47	0.60	0.52	0.63	0.52	0.66
Number of Concurrent Tasks (n=4)	0.47	0.64	0.47	0.64	0.47	0.64
Within Task Demand (maximum value)	0.45	0.62	0.59	0.67	0.49	0.66

Pilot data: continued

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Cumulative VACP (maximum value)	0.45	0.64	0.58	0.69	0.62	0.71
Unacceptable or Marginal Task Conflict (2nd quartile)	0.45	0.26	0.38	0.30	0.40	0.19
W/Index (mean value)	0.41	0.47	0.39	0.47	0.44	0.49
Unacceptable or Marginal Task Conflict (4th quartile)	0.40	0.63	0.38	0.62	0.54	0.62
Cumulative VACP (mean of nonzero values)	0.39	0.50	0.46	0.55	0.68	0.66
Unacceptable or Marginal Task Conflict (1st quartile)	0.39	0.14	0.18	-0.12	0.08	-0.26
Time Occupied Parameter (1st quartile)	0.38	0.28	0.38	0.28	0.38	0.28
Number of Concurrent Tasks (n=3)	0.35	0.07	0.35	0.07	0.35	0.07
Within Task Demand (mean value)	0.35	0.27	0.32	0.28	0.36	0.34
Task Conflict Parameter (1st quartile)	0.32	0.19	-0.10	-0.18	-0.06	-0.24
Unacceptable Task Conflict (1st quartile)	0.32	0.16	0.55	0.42	0.62	0.47
Visual Attentional Demand (mean value)	0.32	0.25	0.32	0.27	0.32	0.28
Cognitive Attentional Demand (mean value)	0.32	0.35	0.45	0.41	0.38	0.40
Task Conflict Parameter (3rd quartile)	0.31	0.28	0.54	0.72	0.45	0.54
Physical Attentional Demand (mean value)	0.28	0.09	0.28	0.07	0.31	0.57
Task Conflict Parameter (4th quartile)	0.26	0.23	0.47	0.40	0.49	0.61
Physical Attentional Demand (mean of nonzero values)	0.21	0.04	0.31	0.00	-0.41	-0.23
Cumulative VACP (mean value)	0.19	0.21	0.27	0.24	0.44	0.48
Physical Attentional Demand (maximum value)	0.18	0.15	0.08	0.16	-0.21	-0.31
Visual Attentional Demand (maximum value)	0.12	0.08	0.39	0.38	0.39	0.39
Time Occupied Parameter (2nd quartile)	0.09	-0.28	0.09	-0.28	0.09	-0.28

Pilot data: continued

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Time Occupied Parameter (3^{rd} quartile)	0.05	-0.18	0.05	-0.18	0.05	-0.18
Task Conflict Parameter (mean value)	0.03	0.22	0.26	0.28	0.42	0.43
Unacceptable or Marginal Task Conflict (mean value)	-0.02	-0.12	0.03	0.09	0.26	0.29
Auditory Attentional Demand (mean of nonzero values)	-0.06	0.02	-0.06	0.03	-0.06	0.03
Across Task Interference (fraction of time zero)	-0.08	0.03	-0.08	0.03	-0.08	0.03
Visual Attentional Demand (fraction of time zero)	-0.11	0.04	-0.11	0.04	-0.11	0.04
Cognitive Attentional Demand (fraction of time zero)	-0.11	0.04	-0.11	0.04	-0.11	0.04
Within Task Demand (fraction of time zero)	-0.11	0.04	-0.11	0.04	-0.11	0.04
W/Index (fraction of time zero)	-0.11	0.04	-0.11	0.04	-0.11	0.04
Cumulative VACP (fraction of time zero)	-0.11	-0.03	-0.11	0.00	-0.11	0.30
Number of Concurrent Tasks (n=1)	-0.12	0.09	-0.12	0.09	-0.12	0.09
Unacceptable Task Conflict (mean value)	-0.14	-0.07	0.02	0.02	-0.21	-0.20
Time Occupied Parameter (mean value)	-0.17	0.02	-0.17	0.03	-0.17	0.03
Physical Attentional Demand (fraction of time zero)	-0.31	-0.07	-0.31	-0.07	-0.40	-0.29
Auditory Attentional Demand (fraction of time zero)	-0.38	-0.31	-0.38	-0.31	-0.38	-0.31
Number of Concurrent Tasks (n=2)	-0.65	-0.44	-0.65	-0.44	-0.65	-0.44

Observer data

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Across Task Interference (mean of nonzero values)	0.52	0.43	0.52	0.43	0.53	0.44
Unacceptable Task Conflict (3rd quartile)	0.51	0.53	0.52	0.50	0.53	0.33
Number of Concurrent Tasks (mean value)	0.49	0.50	0.49	0.50	0.49	0.50
Auditory Attentional Demand (mean value)	0.48	0.45	0.48	0.45	0.48	0.45
Unacceptable or Marginal Task Conflict (mean value)	0.48	0.47	0.47	0.44	0.45	0.45
Number of Concurrent Tasks (maximum value)	0.47	0.45	0.47	0.45	0.47	0.45
Cognitive Attentional Demand (mean of nonzero values)	0.46	0.45	0.33	0.34	0.33	0.34
W/Index (mean of nonzero values)	0.45	0.42	0.46	0.41	0.46	0.42
Time Occupied Parameter (4th quartile)	0.44	0.47	0.44	0.47	0.44	0.47
Within Task Demand (maximum value)	0.43	0.31	0.37	0.34	0.38	0.37
W/Index (mean value)	0.42	0.33	0.42	0.33	0.42	0.33
Cognitive Attentional Demand (maximum value)	0.42	0.35	0.21	0.27	0.21	0.27
Within Task Demand (mean of nonzero values)	0.42	0.45	0.46	0.45	0.48	0.48
Cumulative VACP (mean of nonzero values)	0.42	0.44	0.39	0.40	0.43	0.43
Number of Concurrent Tasks (n=4)	0.41	0.29	0.41	0.29	0.41	0.29
Unacceptable Task Conflict (mean value)	0.41	0.21	0.40	0.38	0.48	0.43
Number of Concurrent Tasks (n=3)	0.40	0.42	0.40	0.42	0.40	0.42
Across Task Interference (mean value)	0.40	0.34	0.40	0.34	0.40	0.34
Number of Concurrent Tasks (n>4)	0.39	0.33	0.39	0.33	0.39	0.33
Unacceptable or Marginal Task Conflict (3rd quartile)	0.37	0.27	0.41	0.28	0.43	0.27

Observer data: continued

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Across Task Interference (maximum value)	0.37	0.25	0.37	0.31	0.37	0.30
W/Index (maximum value)	0.37	0.25	0.33	0.31	0.29	0.32
Task Conflict Parameter (2nd quartile)	0.36	0.22	0.33	0.32	0.30	0.31
Physical Attentional Demand (mean value)	0.33	0.28	0.32	0.27	0.35	0.26
Physical Attentional Demand (mean of nonzero values)	0.33	0.20	0.04	-0.11	0.06	0.08
Time Occupied Parameter (1st quartile)	0.32	-0.01	0.32	-0.01	0.32	-0.01
Visual Attentional Demand (mean of nonzero values)	0.31	0.40	0.42	0.46	0.42	0.45
Cumulative VACP (maximum value)	0.31	0.22	0.39	0.33	0.59	0.55
Visual Attentional Demand (maximum value)	0.30	0.25	0.37	0.36	0.40	0.36
Unacceptable or Marginal Task Conflict (4th quartile)	0.30	0.27	0.39	0.42	0.39	0.42
Physical Attentional Demand (maximum value)	0.30	0.19	0.29	0.15	0.43	0.33
Cognitive Attentional Demand (mean value)	0.30	0.26	0.22	0.22	0.22	0.22
Unacceptable Task Conflict (1st quartile)	0.27	0.05	0.39	0.17	0.36	0.06
Visual Attentional Demand (mean value)	0.25	0.24	0.34	0.29	0.34	0.29
Cumulative VACP (mean value)	0.25	0.22	0.30	0.23	0.24	0.25
Number of Concurrent Tasks (n=2)	0.24	-0.06	0.24	-0.06	0.24	-0.06
Within Task Demand (mean value)	0.23	0.27	0.32	0.28	0.36	0.27
Task Conflict Parameter (mean value)	0.21	0.27	0.41	0.48	0.41	0.45
Unacceptable or Marginal Task Conflict (1st quartile)	0.17	-0.05	0.14	-0.08	0.14	-0.05
Unacceptable or Marginal Task Conflict (2nd quartile)	0.16	-0.02	0.30	0.06	0.30	-0.01
Task Conflict Parameter (1st quartile)	0.16	0.03	0.14	-0.03	0.14	-0.03

Observer data: continued

Workload Measure	Sikorsky Matrices		Combined Matrices		Individual Matrices	
	ρ-value	r-value	ρ-value	r-value	ρ-value	r-value
Time Occupied Parameter (2^{nd} quartile)	0.16	-0.20	0.16	-0.20	0.16	-0.20
Unacceptable Task Conflict (2^{nd} quartile)	0.11	0.24	0.37	0.17	0.41	0.21
Time Occupied Parameter (3^{rd} quartile)	0.09	0.04	0.09	0.04	0.09	0.04
Task Conflict Parameter (3^{rd} quartile)	0.05	0.20	0.37	0.36	0.38	0.35
Task Conflict Parameter (4^{th} quartile)	0.03	0.13	0.41	0.44	0.35	0.42
Auditory Attentional Demand (mean of nonzero values)	0.00	0.02	0.00	0.02	0.00	0.03
Auditory Attentional Demand (maximum value)	0.00	-0.01	0.00	-0.01	0.00	-0.01
Unacceptable Task Conflict (4^{th} quartile)	-0.01	-0.04	0.46	0.46	0.44	0.44
Time Occupied Parameter (mean value)	-0.05	0.06	-0.05	0.06	-0.05	0.06
Visual Attentional Demand (fraction of time zero)	-0.15	-0.07	-0.15	-0.07	-0.15	-0.07
Auditory Attentional Demand (fraction of time zero)	-0.20	-0.08	-0.20	-0.08	-0.20	-0.08
Cognitive Attentional Demand (fraction of time zero)	-0.20	-0.07	-0.20	-0.07	-0.20	-0.07
Within Task Demand (fraction of time zero)	-0.20	-0.07	-0.20	-0.07	-0.20	-0.07
W/Index (fraction of time zero)	-0.20	-0.07	-0.20	-0.07	-0.20	-0.07
Cumulative VACP (fraction of time zero)	-0.20	-0.07	0.22	-0.07	0.16	-0.05
Across Task Interference (fraction of time zero)	-0.26	-0.19	-0.26	-0.19	-0.26	-0.19
Physical Attentional Demand (fraction of time zero)	-0.28	-0.15	-0.28	-0.15	-0.28	-0.14
Number of Concurrent Tasks (n=1)	-0.35	-0.26	-0.35	-0.26	-0.35	-0.26

Appendix B.

Correlation of the empirical workload component measures with both the overall and component analytical workload measures.

The six empirical TLX-like workload component scores were compared to the analytical multi-dimensional workload results, including both at the component values as well as the total score of the W/Index and VACP measures. A non-parametric Spearman rank-order "ρ" correlation was performed between the analytical and empirical data over each of the nine flight segments. Statistical significance testing was done using a 2-tailed t-test and the statistical significance of each correlation is present under the α-columns.

The "Within Task Demand" and the "Across Task Interference" values are the two components of the W/Index relationship of equation 1. The Within Task Demand represents the contributions of each of the tasks to workload assuming no interference; the Across Task Interference represents the conflict associated with processing multiple concurrent tasks. The term parameter "W/Index" represents the sum of these two terms

The "Cognitive Attentional Demand", "Visual Attentional Demand", "Physical Attentional Demand" and "Auditory Attentional Demand" are the four components of the VACP measure. The parameter "Cumulative VACP" represents the sum of the contributions of these four components.

The qualifiers "max", "mean", "mean nonzero" and "zero" indicate the maximum, mean, mean of nonzero values and the incidence of a zero value for a variable over the course of a segment.

Pilots: Individual Conflict Matrix					
Activity Level			**Anxiety Level**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Auditory Attent. Demand mean	0.73	<.0001	Cognitive Attent. Demand max	0.69	<.0001
Cognitive Attent. Demand max	0.71	<.0001	Across Task Interference mean nonzero	0.67	<.0001
Cumulative VACP mean nonzero	0.70	<.0001	Auditory Attent. Demand mean	0.65	<.0001
Across Task Interference mean nonzero	0.69	<.0001	Cumulative VACP mean nonzero	0.65	<.0001
Cumulative VACP max	0.64	<.0001	Across Task Interference max	0.61	<.0001
Across Task Interference max	0.64	<.0001	W/Index max	0.60	<.0001
W/Index max	0.63	<.0001	Cumulative VACP mean	0.57	<.0001
Auditory Attent. Demand max	0.61	<.0001	W/Index mean nonzero	0.56	<.0001
Within Task Demand mean nonzero	0.59	<.0001	Cumulative VACP max	0.56	<.0001
W/Index mean nonzero	0.58	<.0001	Across Task Interference mean	0.54	<.0001
Within Task Demand max	0.51	<.0001	W/Index mean	0.54	<.0001
Visual Attent. Demand mean nonzero	0.51	<.0001	Within Task Demand mean nonzero	0.52	<.0001
Cognitive Attent. Demand mean nonzero	0.50	<.0001	Visual Attent. Demand mean nonzero	0.51	<.0001
Cumulative VACP mean	0.48	<.0001	Cognitive Attent. Demand mean	0.51	<.0001
Across Task Interference mean	0.46	<.0001	Auditory Attent. Demand max	0.50	<.0001
W/Index mean	0.46	<.0001	Within Task Demand max	0.50	<.0001
Cognitive Attent. Demand mean	0.43	<.0001	Within Task Demand mean	0.48	<.0001
Within Task Demand mean	0.43	<.0001	Cognitive Attent. Demand mean nonzero	0.44	<.0001
Visual Attent. Demand max	0.40	0.0012	Visual Attent. Demand mean	0.44	<.0001
Visual Attent. Demand mean	0.35	<.0001	Visual Attent. Demand max	0.43	<.0001
Physical Attent. Demand mean	0.30	0.02	Physical Attent. Demand mean	0.30	0.02
Auditory Attent. Demand mean nonzero	-0.07	0.59	Physical Attent. Demand max	-0.18	0.17
Across Task Interference zero	-0.10	0.42	Auditory Attent. Demand mean nonzero	-0.21	0.10
Visual Attent. Demand zero	-0.15	0.23	Across Task Interference zero	-0.24	0.06
Cognitive Attent. Demand zero	-0.15	0.23	Visual Attent. Demand zero	-0.27	0.03
Within Task Demand zero	-0.15	0.23	Cognitive Attent. Demand zero	-0.27	0.03
W/Index mean zero	-0.15	0.23	Within Task Demand zero	-0.27	0.03
Cumulative VACP #zeros	-0.15	0.23	W/Index mean zero	-0.27	0.03
Physical Attent. Demand max	-0.25	0.05	Cumulative VACP #zeros	-0.27	0.03
Auditory Attent. Demand zero	-0.42	<.0001	Auditory Attent. Demand zero	-0.50	<.0001
Physical Attent. Demand zero	-0.45	<.0001	Physical Attent. Demand zero	-0.52	<.0001
Physical Attent. Demand mean nonzero	-0.46	<.0001	Physical Attent. Demand mean nonzero	-0.54	<.0001

Pilots: Individual Conflict Matrix					
Mental Effort			**Performance**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Across Task Interference mean nonzero	0.74	<.0001	Visual Attent. Demand mean nonzero	0.26	0.04
Across Task Interference max	0.70	<.0001	Visual Attent. Demand max	0.26	0.04
W/Index max	0.68	<.0001	W/Index mean nonzero	0.25	0.05
Cognitive Attent. Demand max	0.68	<.0001	Cognitive Attent. Demand mean	0.21	0.10
Auditory Attent. Demand mean	0.66	<.0001	Within Task Demand mean nonzero	0.19	0.13
W/Index mean nonzero	0.63	<.0001	Visual Attent. Demand mean	0.18	0.15
Cumulative VACP mean nonzero	0.62	<.0001	Across Task Interference mean	0.17	0.19
Cumulative VACP max	0.60	<.0001	W/Index mean	0.17	0.19
Within Task Demand mean nonzero	0.59	<.0001	Within Task Demand max	0.17	0.19
Auditory Attent. Demand max	0.57	<.0001	Across Task Interference mean nonzero	0.16	0.20
Visual Attent. Demand mean nonzero	0.56	<.0001	Within Task Demand mean	0.13	0.31
Within Task Demand max	0.55	<.0001	Cumulative VACP max	0.12	0.35
Visual Attent. Demand max	0.48	<.0001	Across Task Interference max	0.10	0.44
Across Task Interference mean	0.47	<.0001	Cognitive Attent. Demand mean nonzero	0.08	0.53
W/Index mean	0.47	<.0001	Auditory Attent. Demand max	0.08	0.54
Cognitive Attent. Demand mean nonzero	0.42	<.0001	W/Index max	0.06	0.61
Within Task Demand mean	0.41	<.0001	Physical Attent. Demand zero	0.03	0.81
Cumulative VACP mean	0.40	<.0001	Auditory Attent. Demand mean nonzero	0.02	0.85
Cognitive Attent. Demand mean	0.39	<.0001	Cumulative VACP mean	0.01	0.96
Visual Attent. Demand mean	0.37	<.0001	Physical Attent. Demand mean nonzero	0.00	0.97
Physical Attent. Demand mean	0.29	0.02	Physical Attent. Demand max	-0.01	0.92
Auditory Attent. Demand mean nonzero	-0.05	0.71	Auditory Attent. Demand zero	-0.01	0.91
Across Task Interference zero	-0.10	0.45	Auditory Attent. Demand mean	-0.04	0.75
Visual Attent. Demand zero	-0.13	0.33	Cumulative VACP mean nonzero	-0.05	0.67
Cognitive Attent. Demand zero	-0.13	0.33	Visual Attent. Demand zero	-0.06	0.65
Within Task Demand zero	-0.13	0.33	Cognitive Attent. Demand zero	-0.06	0.65
W/Index mean zero	-0.13	0.33	Within Task Demand zero	-0.06	0.65
Cumulative VACP #zeros	-0.13	0.33	W/Index mean zero	-0.06	0.65
Physical Attent. Demand max	-0.18	0.15	Cumulative VACP #zeros	-0.06	0.65
Auditory Attent. Demand zero	-0.37	<.0001	Cognitive Attent. Demand max	-0.06	0.62
Physical Attent. Demand zero	-0.39	<.0001	Across Task Interference zero	-0.11	0.40
Physical Attent. Demand mean nonzero	-0.40	<.0001	Physical Attent. Demand mean	-0.20	0.11

Pilots: Individual Conflict Matrices					
Physical Effort			**Task Complexity**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Across Task Interference mean nonzero	0.56	<.0001	Across Task Interference mean nonzero	0.71	<.0001
Cumulative VACP max	0.55	<.0001	Across Task Interference max	0.68	<.0001
Within Task Demand mean nonzero	0.55	<.0001	Cognitive Attent. Demand max	0.66	<.0001
W/Index mean nonzero	0.53	<.0001	W/Index max	0.66	<.0001
Auditory Attent. Demand max	0.52	<.0001	Auditory Attent. Demand mean	0.63	<.0001
Auditory Attent. Demand mean	0.51	<.0001	W/Index mean nonzero	0.61	<.0001
Visual Attent. Demand mean nonzero	0.49	<.0001	Cumulative VACP mean nonzero	0.60	<.0001
W/Index max	0.48	<.0001	Visual Attent. Demand mean nonzero	0.57	<.0001
Cognitive Attent. Demand max	0.46	<.0001	Cumulative VACP max	0.55	<.0001
Cumulative VACP mean nonzero	0.46	<.0001	Across Task Interference mean	0.54	<.0001
Across Task Interference max	0.44	<.0001	W/Index mean	0.54	<.0001
Within Task Demand max	0.42	<.0001	Within Task Demand mean nonzero	0.54	<.0001
Across Task Interference mean	0.41	<.0001	Within Task Demand max	0.53	<.0001
W/Index mean	0.41	<.0001	Auditory Attent. Demand max	0.51	<.0001
Cognitive Attent. Demand mean nonzero	0.40	<.0001	Cumulative VACP mean	0.50	<.0001
Cognitive Attent. Demand mean	0.37	<.0001	Visual Attent. Demand max	0.50	<.0001
Within Task Demand mean	0.37	<.0001	Within Task Demand mean	0.48	<.0001
Visual Attent. Demand max	0.36	<.0001	Cognitive Attent. Demand mean	0.47	<.0001
Visual Attent. Demand mean	0.33	0.01	Visual Attent. Demand mean	0.45	<.0001
Cumulative VACP mean	0.32	0.01	Cognitive Attent. Demand mean nonzero	0.39	<.0001
Physical Attent. Demand mean	0.02	0.85	Physical Attent. Demand mean	0.32	0.01
Auditory Attent. Demand mean nonzero	-0.03	0.82	Physical Attent. Demand max	-0.16	0.21
Across Task Interference zero	-0.11	0.37	Auditory Attent. Demand mean nonzero	-0.18	0.16
Visual Attent. Demand zero	-0.13	0.32	Across Task Interference zero	-0.21	0.09
Cognitive Attent. Demand zero	-0.13	0.32	Visual Attent. Demand zero	-0.24	0.06
Within Task Demand zero	-0.13	0.32	Cognitive Attent. Demand zero	-0.24	0.06
W/Index mean zero	-0.13	0.32	Within Task Demand zero	-0.24	0.06
Cumulative VACP #zeros	-0.13	0.32	W/Index mean zero	-0.24	0.06
Physical Attent. Demand max	-0.28	0.03	Cumulative VACP #zeros	-0.24	0.06
Physical Attent. Demand zero	-0.32	0.01	Auditory Attent. Demand zero	-0.45	<.0001
Auditory Attent. Demand zero	-0.33	0.01	Physical Attent. Demand zero	-0.48	<.0001
Physical Attent. Demand mean nonzero	-0.33	0.01	Physical Attent. Demand mean nonzero	-0.50	<.0001

Pilots: Combined Conflict Matrix					
Activity Level			Anxiety Level		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Auditory Attent. Demand mean	0.73	<.0001	Across Task Interference mean nonzero	0.67	<.0001
Across Task Interference mean nonzero	0.69	<.0001	W/Index max	0.66	<.0001
W/Index max	0.68	<.0001	Auditory Attent. Demand mean	0.65	<.0001
Across Task Interference max	0.67	<.0001	Within Task Demand max	0.65	<.0001
Cognitive Attent. Demand max	0.66	<.0001	Across Task Interference max	0.65	<.0001
Cumulative VACP max	0.62	<.0001	Cognitive Attent. Demand max	0.62	<.0001
Within Task Demand max	0.61	<.0001	Across Task Interference mean	0.61	<.0001
Auditory Attent. Demand max	0.61	<.0001	Cumulative VACP max	0.57	<.0001
W/Index mean nonzero	0.58	<.0001	W/Index mean nonzero	0.56	<.0001
Across Task Interference mean	0.54	<.0001	Cognitive Attent. Demand mean	0.56	<.0001
Visual Attent. Demand mean nonzero	0.51	<.0001	Visual Attent. Demand mean nonzero	0.51	<.0001
Cognitive Attent. Demand mean nonzero	0.50	<.0001	W/Index mean	0.50	<.0001
Within Task Demand mean nonzero	0.50	<.0001	Auditory Attent. Demand max	0.50	<.0001
Cognitive Attent. Demand mean	0.50	<.0001	Within Task Demand mean nonzero	0.48	<.0001
Cumulative VACP mean nonzero	0.49	<.0001	Visual Attent. Demand max	0.48	<.0001
W/Index mean	0.42	<.0001	Cognitive Attent. Demand mean nonzero	0.44	0.03
Visual Attent. Demand max	0.42	<.0001	Visual Attent. Demand mean	0.44	0.0004
Visual Attent. Demand mean	0.35	<.0001	Within Task Demand mean	0.44	0.0004
Within Task Demand mean	0.35	<.0001	Cumulative VACP mean	0.41	0.0008
Cumulative VACP mean	0.34	0.01	Physical Attent. Demand mean	0.41	0.0008
Physical Attent. Demand mean nonzero	0.32	0.01	Cumulative VACP mean nonzero	0.38	0.0021
Physical Attent. Demand mean	0.31	0.01	Physical Attent. Demand mean nonzero	0.28	0.0285
Physical Attent. Demand max	0.02	0.90	Physical Attent. Demand max	0.11	0.3719
Auditory Attent. Demand mean nonzero	-0.07	0.59	Auditory Attent. Demand mean nonzero	-0.21	0.1001
Across Task Interference zero	-0.10	0.42	Across Task Interference zero	-0.24	0.0636
Visual Attent. Demand zero	-0.15	0.23	Visual Attent. Demand zero	-0.27	0.0332
Cognitive Attent. Demand zero	-0.15	0.23	Cognitive Attent. Demand zero	-0.27	0.0332
Within Task Demand zero	-0.15	0.23	Within Task Demand zero	-0.27	0.0332
W/Index zero	-0.15	0.23	W/Index zero	-0.278	0.0332
Cumulative VACP #zeros	-0.15	0.23	Cumulative VACP #zeros	-0.27	0.0332
Physical Attent. Demand zero	-0.34	0.01	Physical Attent. Demand zero	-0.42	0.0006
Auditory Attent. Demand zero	-0.42	<.0001	Auditory Attent. Demand zero	-0.502	<.0001

Pilots: Combined Conflict Matrix					
Mental Effort			**Performance**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Across Task Interference mean nonzero	0.74	<.0001	Visual Attent. Demand mean nonzero	0.26	0.04
Across Task Interference max	0.72	<.0001	Within Task Demand mean nonzero	0.26	0.04
W/Index max	0.70	<.0001	Visual Attent. Demand max	0.26	0.04
Auditory Attent. Demand mean	0.66	<.0001	W/Index mean nonzero	0.25	0.05
W/Index mean nonzero	0.63	<.0001	Physical Attent. Demand mean	0.23	0.06
Cognitive Attent. Demand max	0.62	<.0001	Cumulative VACP max	0.23	0.07
Within Task Demand max	0.61	<.0001	Physical Attent. Demand max	0.22	0.09
Cumulative VACP max	0.60	<.0001	Cognitive Attent. Demand mean	0.20	0.12
Within Task Demand mean nonzero	0.57	<.0001	Within Task Demand max	0.19	0.13
Auditory Attent. Demand max	0.57	<.0001	Cumulative VACP mean nonzero	0.19	0.14
Visual Attent. Demand mean nonzero	0.56	<.0001	Visual Attent. Demand mean	0.18	0.15
Across Task Interference mean	0.55	<.0001	Within Task Demand mean	0.18	0.15
Cognitive Attent. Demand mean	0.47	<.0001	W/Index mean	0.17	0.18
Cumulative VACP mean nonzero	0.46	<.0001	Across Task Interference mean nonzero	0.16	0.20
Visual Attent. Demand max	0.44	<.0001	Across Task Interference mean	0.16	0.21
W/Index mean	0.44	<.0001	Across Task Interference max	0.13	0.31
Cognitive Attent. Demand mean nonzero	0.42	<.0001	Cumulative VACP mean	0.10	0.43
Visual Attent. Demand mean	0.37	<.0001	Cognitive Attent. Demand mean nonzero	0.08	0.53
Within Task Demand mean	0.37	<.0001	Auditory Attent. Demand max	0.08	0.54
Physical Attent. Demand mean	0.32	0.01	W/Index max	0.05	0.67
Physical Attent. Demand mean nonzero	0.32	0.01	Auditory Attent. Demand mean nonzero	0.02	0.85
Cumulative VACP mean	0.31	0.01	Auditory Attent. Demand zero	-0.01	0.91
Physical Attent. Demand max	0.11	0.38	Physical Attent. Demand mean nonzero	-0.02	0.87
Auditory Attent. Demand mean nonzero	-0.05	0.71	Auditory Attent. Demand mean	-0.04	0.75
Across Task Interference zero	-0.10	0.45	Visual Attent. Demand zero	-0.06	0.65
Visual Attent. Demand zero	-0.13	0.33	Cognitive Attent. Demand zero	-0.06	0.65
Cognitive Attent. Demand zero	-0.13	0.33	Within Task Demand zero	-0.06	0.65
Within Task Demand zero	-0.13	0.33	W/Index zero	-0.06	0.65
W/Index zero	-0.13	0.33	Cumulative VACP #zeros	-0.06	0.65
Cumulative VACP #zeros	-0.13	0.33	Cognitive Attent. Demand max	-0.10	0.44
Physical Attent. Demand zero	-0.36	<.0001	Across Task Interference zero	-0.11	0.40
Auditory Attent. Demand zero	-0.37	<.0001	Physical Attent. Demand zero	-0.15	0.25

Pilots: Combined Conflict Matrix					
Physical Effort			**Task Complexity**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Across Task Interference mean nonzero	0.56	<.0001	Across Task Interference mean nonzero	0.71	<.0001
Cumulative VACP max	0.56	<.0001	Across Task Interference max	0.71	<.0001
W/Index mean nonzero	0.53	<.0001	W/Index max	0.70	<.0001
Auditory Attent. Demand max	0.52	<.0001	Within Task Demand max	0.64	<.0001
Auditory Attent. Demand mean	0.51	<.0001	Auditory Attent. Demand mean	0.63	<.0001
Visual Attent. Demand mean nonzero	0.49	<.0001	W/Index mean nonzero	0.61	<.0001
Across Task Interference max	0.48	<.0001	Across Task Interference mean	0.61	<.0001
Within Task Demand mean nonzero	0.47	<.0001	Cognitive Attent. Demand max	0.59	<.0001
Within Task Demand max	0.47	<.0001	Cumulative VACP max	0.58	<.0001
Cumulative VACP mean nonzero	0.46	<.0001	Visual Attent. Demand mean nonzero	0.57	<.0001
Across Task Interference mean	0.46	<.0001	Cognitive Attent. Demand mean	0.53	<.0001
W/Index max	0.45	<.0001	Within Task Demand mean nonzero	0.53	<.0001
Cognitive Attent. Demand mean	0.43	<.0001	W/Index mean	0.52	<.0001
Cognitive Attent. Demand max	0.42	<.0001	Auditory Attent. Demand max	0.51	<.0001
Visual Attent. Demand max	0.40	<.0001	Visual Attent. Demand max	0.49	<.0001
Cognitive Attent. Demand mean nonzero	0.40	<.0001	Visual Attent. Demand mean	0.45	<.0001
W/Index mean	0.37	<.0001	Within Task Demand mean	0.45	<.0001
Visual Attent. Demand mean	0.33	0.01	Physical Attent. Demand mean	0.41	<.0001
Within Task Demand mean	0.33	0.01	Cumulative VACP mean	0.39	<.0001
Cumulative VACP mean	0.29	0.02	Cumulative VACP mean nonzero	0.39	<.0001
Physical Attent. Demand mean	0.28	0.02	Cognitive Attent. Demand mean nonzero	0.39	<.0001
Physical Attent. Demand mean nonzero	0.15	0.25	Physical Attent. Demand mean nonzero	0.27	0.03
Physical Attent. Demand max	0.05	0.72	Physical Attent. Demand max	0.13	0.33
Auditory Attent. Demand mean nonzero	-0.03	0.82	Auditory Attent. Demand mean nonzero	-0.18	0.16
Across Task Interference zero	-0.11	0.37	Across Task Interference zero	-0.21	0.09
Visual Attent. Demand zero	-0.13	0.32	Visual Attent. Demand zero	-0.24	0.06
Cognitive Attent. Demand zero	-0.13	0.32	Cognitive Attent. Demand zero	-0.24	0.06
Within Task Demand zero	-0.13	0.32	Within Task Demand zero	-0.24	0.06
W/Index zero	-0.13	0.32	W/Index zero	-0.24	0.06
Cumulative VACP #zeros	-0.13	0.32	Cumulative VACP #zeros	-0.24	0.06
Physical Attent. Demand zero	-0.32	0.01	Physical Attent. Demand zero	-0.44	<.0001
Auditory Attent. Demand zero	-0.33	0.01	Auditory Attent. Demand zero	-0.45	<.0001

Pilots: Sikorsky Conflict Matrix					
Activity Level			Anxiety Level		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Within Task Demand mean nonzero	0.71	<.0001	Within Task Demand mean nonzero	0.64	<.0001
Cumulative VACP mean nonzero	0.71	<.0001	Cumulative VACP mean nonzero	0.64	<.0001
Cognitive Attent. Demand mean nonzero	0.69	<.0001	Cognitive Attent. Demand mean nonzero	0.63	<.0001
Cumulative VACP max	0.65	<.0001	Cumulative VACP max	0.59	<.0001
W/Index mean nonzero	0.63	<.0001	W/Index mean nonzero	0.58	<.0001
Across Task Interference mean nonzero	0.60	<.0001	Visual Attent. Demand max	0.56	<.0001
Visual Attent. Demand max	0.58	<.0001	Visual Attent. Demand mean nonzero	0.56	<.0001
Visual Attent. Demand mean nonzero	0.58	<.0001	Across Task Interference mean nonzero	0.55	<.0001
Within Task Demand max	0.55	<.0001	Within Task Demand max	0.52	<.0001
Across Task Interference max	0.51	<.0001	Physical Attent. Demand mean	0.51	<.0001
W/Index max	0.51	<.0001	Across Task Interference max	0.48	<.0001
Physical Attent. Demand mean	0.50	<.0001	W/Index max	0.48	<.0001
W/Index mean	0.47	<.0001	W/Index mean	0.48	<.0001
Physical Attent. Demand max	0.44	0.01	Physical Attent. Demand max	0.46	<.0001
Cognitive Attent. Demand max	0.42	0.01	Auditory Attent. Demand mean	0.42	0.01
Auditory Attent. Demand mean	0.41	0.01	Cognitive Attent. Demand max	0.42	0.01
Across Task Interference mean	0.41	0.01	Across Task Interference mean	0.42	0.01
Visual Attent. Demand mean	0.31	0.07	Visual Attent. Demand mean	0.35	0.04
Cumulative VACP mean	0.31	0.07	Cumulative VACP mean	0.35	0.04
Within Task Demand mean	0.31	0.07	Within Task Demand mean	0.35	0.04
Physical Attent. Demand mean nonzero	0.29	0.09	Cognitive Attent. Demand mean	0.28	0.09
Cognitive Attent. Demand mean	0.23	0.17	Physical Attent. Demand mean nonzero	0.24	0.15
Visual Attent. Demand zero	0.03	0.85	Auditory Attent. Demand max	0.01	0.95
Auditory Attent. Demand max	-0.03	0.85	Visual Attent. Demand zero	-0.08	0.64
Cognitive Attent. Demand zero	-0.06	0.71	Cognitive Attent. Demand zero	-0.16	0.34
Within Task Demand zero	-0.06	0.71	Within Task Demand zero	-0.16	0.34
W/Index mean zero	-0.06	0.71	W/Index mean zero	-0.16	0.34
Cumulative VACP #zeros	-0.06	0.71	Cumulative VACP #zeros	-0.16	0.34
Auditory Attent. Demand mean nonzero	-0.16	0.34	Auditory Attent. Demand mean nonzero	-0.24	0.16
Auditory Attent. Demand zero	-0.36	0.03	Auditory Attent. Demand zero	-0.40	0.01
Across Task Interference zero	-0.38	0.02	Across Task Interference zero	-0.40	0.01
Physical Attent. Demand zero	-0.44	0.01	Physical Attent. Demand zero	-0.46	<.0001

Pilots Sikorsky Conflict Matrix:					
Mental Effort			**Performance**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Within Task Demand mean nonzero	0.43	0.01	Cognitive Attent. Demand mean	0.06	0.74
Cumulative VACP mean nonzero	0.43	0.01	Auditory Attent. Demand zero	0.04	0.81
Cognitive Attent. Demand mean nonzero	0.43	0.01	Physical Attent. Demand mean nonzero	0.03	0.85
Cumulative VACP max	0.39	0.02	Auditory Attent. Demand max	0.03	0.86
W/Index mean nonzero	0.38	0.02	Visual Attent. Demand mean	0.02	0.89
Physical Attent. Demand mean	0.36	0.03	Cumulative VACP mean	0.02	0.89
Visual Attent. Demand mean nonzero	0.36	0.03	Physical Attent. Demand zero	0.02	0.89
Across Task Interference mean nonzero	0.35	0.03	Auditory Attent. Demand mean nonzero	0.02	0.90
Visual Attent. Demand max	0.35	0.04	Across Task Interference mean	0.02	0.91
Within Task Demand max	0.32	0.06	Auditory Attent. Demand mean	0.02	0.92
W/Index mean	0.30	0.07	Across Task Interference zero	0.00	0.98
Physical Attent. Demand max	0.30	0.08	Within Task Demand mean	0.00	0.98
Across Task Interference max	0.29	0.09	W/Index mean	0.00	0.99
W/Index max	0.29	0.09	Cognitive Attent. Demand max	0.00	1.00
Auditory Attent. Demand mean	0.25	0.13	Physical Attent. Demand mean	-0.01	0.95
Across Task Interference mean	0.25	0.14	Physical Attent. Demand max	-0.04	0.83
Cognitive Attent. Demand max	0.24	0.16	Across Task Interference mean nonzero	-0.04	0.80
Physical Attent. Demand mean nonzero	0.24	0.16	Within Task Demand max	-0.06	0.74
Visual Attent. Demand mean	0.24	0.16	W/Index mean nonzero	-0.06	0.71
Cumulative VACP mean	0.24	0.16	Across Task Interference max	-0.08	0.65
Within Task Demand mean	0.21	0.21	W/Index max	-0.08	0.65
Cognitive Attent. Demand mean	0.16	0.35	Cognitive Attent. Demand mean nonzero	-0.08	0.64
Auditory Attent. Demand max	0.03	0.87	Visual Attent. Demand mean nonzero	-0.08	0.62
Visual Attent. Demand zero	-0.05	0.79	Cognitive Attent. Demand zero	-0.10	0.56
Cognitive Attent. Demand zero	-0.09	0.61	Within Task Demand zero	-0.10	0.56
Within Task Demand zero	-0.09	0.61	W/Index mean zero	-0.10	0.56
W/Index mean zero	-0.09	0.61	Cumulative VACP #zeros	-0.10	0.56
Cumulative VACP #zeros	-0.09	0.61	Visual Attent. Demand max	-0.12	0.49
Auditory Attent. Demand mean nonzero	-0.12	0.47	Within Task Demand mean nonzero	-0.13	0.46
Across Task Interference zero	-0.24	0.15	Cumulative VACP mean nonzero	-0.13	0.46
Auditory Attent. Demand zero	-0.25	0.14	Visual Attent. Demand zero	-0.14	0.41
Physical Attent. Demand zero	-0.30	0.08	Cumulative VACP max	-0.15	0.38

Pilots: Sikorsky Conflict Matrix					
Physical Effort			**Task Complexity**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Within Task Demand mean nonzero	0.39	0.02	Within Task Demand mean nonzero	0.70	<.0001
Cumulative VACP mean nonzero	0.39	0.02	Cumulative VACP mean nonzero	0.70	<.0001
Cognitive Attent. Demand mean nonzero	0.35	0.04	Cognitive Attent. Demand mean nonzero	0.67	<.0001
Cumulative VACP max	0.35	0.04	Cumulative VACP max	0.66	<.0001
W/Index mean nonzero	0.33	0.05	W/Index mean nonzero	0.64	<.0001
Across Task Interference mean nonzero	0.33	0.05	Visual Attent. Demand mean nonzero	0.61	<.0001
Across Task Interference max	0.32	0.06	Visual Attent. Demand max	0.60	<.0001
W/Index max	0.32	0.06	Across Task Interference mean nonzero	0.59	<.0001
Within Task Demand max	0.31	0.07	Within Task Demand max	0.58	<.0001
Visual Attent. Demand max	0.31	0.07	Across Task Interference max	0.56	<.0001
Visual Attent. Demand mean nonzero	0.28	0.10	W/Index max	0.56	<.0001
Auditory Attent. Demand mean	0.24	0.15	Physical Attent. Demand mean	0.53	<.0001
Cognitive Attent. Demand max	0.23	0.17	W/Index mean	0.51	<.0001
W/Index mean	0.22	0.20	Physical Attent. Demand max	0.47	<.0001
Across Task Interference mean	0.19	0.26	Cognitive Attent. Demand max	0.47	<.0001
Physical Attent. Demand mean	0.19	0.27	Across Task Interference mean	0.46	0.01
Physical Attent. Demand max	0.17	0.33	Auditory Attent. Demand mean	0.43	0.01
Within Task Demand mean	0.15	0.39	Within Task Demand mean	0.40	0.02
Cognitive Attent. Demand mean	0.13	0.45	Visual Attent. Demand mean	0.38	0.02
Visual Attent. Demand mean	0.12	0.50	Cumulative VACP mean	0.38	0.02
Cumulative VACP mean	0.12	0.50	Cognitive Attent. Demand mean	0.33	0.05
Physical Attent. Demand mean nonzero	0.09	0.59	Physical Attent. Demand mean nonzero	0.21	0.21
Visual Attent. Demand zero	0.08	0.63	Auditory Attent. Demand max	0.01	0.96
Auditory Attent. Demand max	0.05	0.78	Visual Attent. Demand zero	-0.05	0.76
Cognitive Attent. Demand zero	-0.03	0.88	Cognitive Attent. Demand zero	-0.17	0.33
Within Task Demand zero	-0.03	0.88	Within Task Demand zero	-0.17	0.33
W/Index mean zero	-0.03	0.88	W/Index mean zero	-0.17	0.33
Cumulative VACP #zeros	-0.03	0.88	Cumulative VACP #zeros	-0.17	0.33
Auditory Attent. Demand mean nonzero	-0.03	0.85	Auditory Attent. Demand mean nonzero	-0.27	0.12
Auditory Attent. Demand zero	-0.16	0.36	Across Task Interference zero	-0.44	0.01
Across Task Interference zero	-0.16	0.36	Auditory Attent. Demand zero	-0.44	0.01
Physical Attent. Demand zero	-0.18	0.29	Physical Attent. Demand zero	-0.49	<.0001

Observers: Individual Conflict Matrix					
Activity Level			**Anxiety Level**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Cumulative VACP mean nonzero	0.73	<.0001	Cognitive Attent. Demand mean nonzero	0.62	<.0001
Cognitive Attent. Demand mean nonzero	0.72	<.0001	Cumulative VACP mean nonzero	0.62	<.0001
Within Task Demand mean nonzero	0.58	<.0001	Within Task Demand mean nonzero	0.54	<.0001
Across Task Interference mean nonzero	0.53	<.0001	Across Task Interference mean nonzero	0.51	<.0001
W/Index mean nonzero	0.52	<.0001	W/Index mean nonzero	0.49	<.0001
Across Task Interference max	0.51	<.0001	Across Task Interference max	0.48	<.0001
Physical Attent. Demand max	0.49	<.0001	W/Index mean	0.48	<.0001
W/Index mean	0.47	<.0001	Auditory Attent. Demand mean	0.42	0.01
Auditory Attent. Demand mean	0.41	0.01	Across Task Interference mean	0.42	0.01
Across Task Interference mean	0.41	0.01	Physical Attent. Demand mean	0.41	0.01
Visual Attent. Demand mean nonzero	0.38	0.02	Physical Attent. Demand max	0.41	0.01
Cognitive Attent. Demand mean	0.37	0.03	Cognitive Attent. Demand mean	0.40	0.01
Visual Attent. Demand max	0.36	0.03	Visual Attent. Demand mean nonzero	0.36	0.03
Physical Attent. Demand mean	0.36	0.03	Visual Attent. Demand max	0.35	0.03
W/Index max	0.34	0.04	Cumulative VACP max	0.34	0.04
Cumulative VACP max	0.32	0.06	W/Index max	0.34	0.05
Within Task Demand max	0.26	0.13	Within Task Demand max	0.28	0.09
Cognitive Attent. Demand max	0.24	0.16	Cumulative VACP mean	0.28	0.09
Cumulative VACP mean	0.23	0.17	Cognitive Attent. Demand max	0.27	0.11
Within Task Demand mean	0.21	0.22	Within Task Demand mean	0.27	0.12
Visual Attent. Demand mean	0.09	0.60	Visual Attent. Demand mean	0.16	0.36
Visual Attent. Demand zero	0.03	0.85	Auditory Attent. Demand max	0.01	0.95
Physical Attent. Demand mean nonzero	0.03	0.88	Physical Attent. Demand mean nonzero	-0.03	0.88
Auditory Attent. Demand max	-0.03	0.85	Visual Attent. Demand zero	-0.08	0.64
Cognitive Attent. Demand zero	-0.06	0.71	Cognitive Attent. Demand zero	-0.16	0.34
Within Task Demand zero	-0.06	0.71	Within Task Demand zero	-0.16	0.34
W/Index mean zero	-0.06	0.71	W/Index mean zero	-0.16	0.34
Cumulative VACP #zeros	-0.06	0.71	Cumulative VACP #zeros	-0.16	0.34
Auditory Attent. Demand mean nonzero	-0.16	0.34	Auditory Attent. Demand mean nonzero	-0.24	0.16
Auditory Attent. Demand zero	-0.36	0.03	Auditory Attent. Demand zero	-0.40	0.01
Across Task Interference zero	-0.38	0.02	Across Task Interference zero	-0.40	0.01
Physical Attent. Demand zero	-0.44	0.01	Physical Attent. Demand zero	-0.46	<.0001

Observers: Individual Conflict Matrix					
Mental Effort			**Performance**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Cumulative VACP mean nonzero	0.44	0.01	Visual Attent. Demand mean	0.13	0.46
Cognitive Attent. Demand mean nonzero	0.44	0.01	Within Task Demand mean	0.07	0.68
Within Task Demand mean nonzero	0.34	0.04	Cumulative VACP mean	0.06	0.74
Across Task Interference mean nonzero	0.33	0.05	Physical Attent. Demand mean	0.05	0.76
W/Index mean	0.30	0.07	Within Task Demand max	0.05	0.79
W/Index mean nonzero	0.29	0.09	Auditory Attent. Demand zero	0.04	0.81
Across Task Interference max	0.29	0.09	Auditory Attent. Demand max	0.03	0.86
Physical Attent. Demand mean	0.28	0.10	Cumulative VACP max	0.03	0.87
Cognitive Attent. Demand mean	0.26	0.13	Physical Attent. Demand zero	0.02	0.89
Auditory Attent. Demand mean	0.25	0.13	Auditory Attent. Demand mean nonzero	0.02	0.90
Across Task Interference mean	0.25	0.14	Across Task Interference mean	0.02	0.91
Physical Attent. Demand max	0.24	0.16	Auditory Attent. Demand mean	0.02	0.92
Visual Attent. Demand mean nonzero	0.19	0.27	Visual Attent. Demand mean nonzero	0.01	0.94
Visual Attent. Demand max	0.19	0.28	Visual Attent. Demand max	0.01	0.95
Cumulative VACP max	0.19	0.28	Cognitive Attent. Demand max	0.01	0.96
W/Index max	0.17	0.34	Across Task Interference zero	0.00	0.98
Cumulative VACP mean	0.16	0.35	W/Index mean	0.00	0.99
Cognitive Attent. Demand max	0.15	0.39	Across Task Interference mean nonzero	-0.01	0.97
Within Task Demand mean	0.14	0.43	Physical Attent. Demand mean nonzero	-0.02	0.90
Within Task Demand max	0.13	0.43	Cognitive Attent. Demand mean	-0.02	0.89
Visual Attent. Demand mean	0.06	0.75	W/Index mean nonzero	-0.03	0.88
Auditory Attent. Demand max	0.03	0.87	W/Index max	-0.04	0.80
Physical Attent. Demand mean nonzero	0.02	0.90	Within Task Demand mean nonzero	-0.04	0.80
Visual Attent. Demand zero	-0.05	0.79	Physical Attent. Demand max	-0.07	0.67
Cognitive Attent. Demand zero	-0.09	0.61	Across Task Interference max	-0.08	0.65
Within Task Demand zero	-0.09	0.61	Cognitive Attent. Demand zero	-0.10	0.56
W/Index mean zero	-0.09	0.61	Within Task Demand zero	-0.10	0.56
Cumulative VACP #zeros	-0.09	0.61	W/Index mean zero	-0.10	0.56
Auditory Attent. Demand mean nonzero	-0.12	0.47	Cumulative VACP #zeros	-0.10	0.56
Across Task Interference zero	-0.24	0.15	Cumulative VACP mean nonzero	-0.13	0.46
Auditory Attent. Demand zero	-0.25	0.14	Visual Attent. Demand zero	-0.14	0.41
Physical Attent. Demand zero	-0.30	0.08	Cognitive Attent. Demand mean nonzero	-0.15	0.38

Observers: Individual Conflict Matrix					
Physical Effort			**Task Complexity**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Cumulative VACP mean nonzero	0.37	0.03	Cognitive Attent. Demand mean nonzero	0.68	<.0001
Cognitive Attent. Demand mean nonzero	0.37	0.03	Cumulative VACP mean nonzero	0.66	<.0001
Physical Attent. Demand max	0.36	0.03	Within Task Demand mean nonzero	0.60	<.0001
Within Task Demand mean nonzero	0.33	0.05	Across Task Interference max	0.56	<.0001
Across Task Interference max	0.32	0.06	W/Index mean nonzero	0.54	<.0001
W/Index mean nonzero	0.30	0.08	Across Task Interference mean nonzero	0.54	<.0001
Across Task Interference mean nonzero	0.29	0.09	W/Index mean	0.51	<.0001
W/Index max	0.25	0.15	Physical Attent. Demand max	0.49	<.0001
Auditory Attent. Demand mean	0.24	0.15	Across Task Interference mean	0.46	0.01
W/Index mean	0.22	0.20	Cognitive Attent. Demand mean	0.45	0.01
Visual Attent. Demand mean nonzero	0.22	0.20	Auditory Attent. Demand mean	0.43	0.01
Cumulative VACP max	0.21	0.23	W/Index max	0.42	0.01
Visual Attent. Demand max	0.21	0.23	Visual Attent. Demand mean nonzero	0.41	0.01
Across Task Interference mean	0.19	0.26	Visual Attent. Demand max	0.40	0.01
Cognitive Attent. Demand max	0.18	0.29	Cumulative VACP max	0.40	0.02
Within Task Demand max	0.18	0.29	Physical Attent. Demand mean	0.38	0.02
Cognitive Attent. Demand mean	0.16	0.34	Cognitive Attent. Demand max	0.35	0.03
Physical Attent. Demand mean	0.14	0.41	Within Task Demand max	0.34	0.04
Within Task Demand mean	0.14	0.43	Cumulative VACP mean	0.33	0.05
Cumulative VACP mean	0.13	0.45	Within Task Demand mean	0.30	0.08
Visual Attent. Demand zero	0.08	0.63	Visual Attent. Demand mean	0.18	0.31
Physical Attent. Demand mean nonzero	0.07	0.70	Auditory Attent. Demand max	0.01	0.96
Visual Attent. Demand mean	0.05	0.77	Visual Attent. Demand zero	-0.05	0.76
Auditory Attent. Demand max	0.05	0.78	Physical Attent. Demand mean nonzero	-0.08	0.65
Cognitive Attent. Demand zero	-0.03	0.88	Cognitive Attent. Demand zero	-0.17	0.33
Within Task Demand zero	-0.03	0.88	Within Task Demand zero	-0.17	0.33
W/Index mean zero	-0.03	0.88	W/Index mean zero	-0.17	0.33
Cumulative VACP #zeros	-0.03	0.88	Cumulative VACP #zeros	-0.17	0.33
Auditory Attent. Demand mean nonzero	-0.03	0.85	Auditory Attent. Demand mean nonzero	-0.27	0.12
Auditory Attent. Demand zero	-0.16	0.36	Across Task Interference zero	-0.44	0.01
Across Task Interference zero	-0.16	0.36	Auditory Attent. Demand zero	-0.44	0.01
Physical Attent. Demand zero	-0.18	0.29	Physical Attent. Demand zero	-0.49	<.0001

Observers: Combined Conflict Matrix					
Activity Level			Anxiety Level		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Cognitive Attent. Demand mean nonzero	0.72	<.0001	Within Task Demand mean nonzero	0.63	<.0001
Cumulative VACP mean nonzero	0.71	<.0001	Cognitive Attent. Demand mean nonzero	0.62	<.0001
Within Task Demand mean nonzero	0.69	<.0001	Cumulative VACP mean nonzero	0.61	<.0001
Across Task Interference mean nonzero	0.60	<.0001	Physical Attent. Demand max	0.55	<.0001
Physical Attent. Demand max	0.58	<.0001	Across Task Interference mean nonzero	0.55	<.0001
W/Index mean nonzero	0.52	<.0001	Physical Attent. Demand mean	0.51	<.0001
Across Task Interference max	0.51	<.0001	W/Index mean nonzero	0.49	<.0001
Physical Attent. Demand mean	0.50	<.0001	Across Task Interference max	0.48	<.0001
W/Index mean	0.47	<.0001	W/Index mean	0.48	<.0001
W/Index max	0.44	0.01	Auditory Attent. Demand mean	0.42	0.01
Auditory Attent. Demand mean	0.41	0.01	Across Task Interference mean	0.42	0.01
Across Task Interference mean	0.41	0.01	W/Index max	0.42	0.01
Cumulative VACP max	0.39	0.02	Cognitive Attent. Demand mean	0.40	0.01
Visual Attent. Demand mean nonzero	0.38	0.02	Cumulative VACP max	0.40	0.02
Cognitive Attent. Demand mean	0.37	0.03	Visual Attent. Demand mean nonzero	0.36	0.03
Within Task Demand max	0.33	0.05	Within Task Demand max	0.34	0.04
Within Task Demand mean	0.25	0.13	Within Task Demand mean	0.30	0.08
Cognitive Attent. Demand max	0.24	0.16	Cumulative VACP mean	0.28	0.09
Cumulative VACP mean	0.23	0.17	Cognitive Attent. Demand max	0.27	0.11
Visual Attent. Demand max	0.20	0.25	Visual Attent. Demand max	0.22	0.21
Physical Attent. Demand mean nonzero	0.13	0.47	Visual Attent. Demand mean	0.16	0.36
Visual Attent. Demand mean	0.09	0.60	Physical Attent. Demand mean nonzero	0.11	0.52
Visual Attent. Demand zero	0.03	0.85	Auditory Attent. Demand max	0.01	0.95
Auditory Attent. Demand max	-0.03	0.85	Visual Attent. Demand zero	-0.08	0.64
Cognitive Attent. Demand zero	-0.06	0.71	Cognitive Attent. Demand zero	-0.16	0.34
Within Task Demand zero	-0.06	0.71	Within Task Demand zero	-0.16	0.34
W/Index mean zero	-0.06	0.71	W/Index mean zero	-0.16	0.34
Cumulative VACP #zeros	-0.06	0.71	Cumulative VACP #zeros	-0.16	0.34
Auditory Attent. Demand mean nonzero	-0.16	0.34	Auditory Attent. Demand mean nonzero	-0.24	0.16
Auditory Attent. Demand zero	-0.36	0.03	Auditory Attent. Demand zero	-0.40	0.01
Across Task Interference zero	-0.38	0.02	Across Task Interference zero	-0.40	0.01
Physical Attent. Demand zero	-0.44	0.01	Physical Attent. Demand zero	-0.46	<.0001

Observers: Combined Conflict Matrix					
Mental Effort			**Performance**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Cognitive Attent. Demand mean nonzero	0.44	0.01	Visual Attent. Demand mean	0.13	0.46
Within Task Demand mean nonzero	0.43	0.01	Cumulative VACP mean	0.06	0.74
Cumulative VACP mean nonzero	0.43	0.01	Visual Attent. Demand max	0.06	0.74
Across Task Interference mean nonzero	0.35	0.03	Within Task Demand mean	0.05	0.78
Physical Attent. Demand max	0.34	0.04	Auditory Attent. Demand zero	0.04	0.81
Physical Attent. Demand mean	0.34	0.04	Auditory Attent. Demand max	0.03	0.86
W/Index mean	0.30	0.07	Physical Attent. Demand zero	0.02	0.89
W/Index mean nonzero	0.29	0.09	Auditory Attent. Demand mean nonzero	0.02	0.90
Across Task Interference max	0.29	0.09	Across Task Interference mean	0.02	0.91
Cognitive Attent. Demand mean	0.26	0.13	Within Task Demand max	0.02	0.92
Auditory Attent. Demand mean	0.25	0.13	Auditory Attent. Demand mean	0.02	0.92
Across Task Interference mean	0.25	0.14	Visual Attent. Demand mean nonzero	0.01	0.94
Cumulative VACP max	0.23	0.18	Cognitive Attent. Demand max	0.01	0.96
W/Index max	0.23	0.18	Across Task Interference zero	0.00	0.98
Visual Attent. Demand mean nonzero	0.19	0.27	W/Index mean	0.00	0.99
Within Task Demand max	0.18	0.30	Cumulative VACP max	0.00	1.00
Within Task Demand mean	0.17	0.33	Cognitive Attent. Demand mean	-0.02	0.89
Cumulative VACP mean	0.16	0.35	W/Index mean nonzero	-0.03	0.88
Cognitive Attent. Demand max	0.15	0.39	Physical Attent. Demand mean	-0.03	0.86
Physical Attent. Demand mean nonzero	0.09	0.61	Physical Attent. Demand mean nonzero	-0.03	0.84
Visual Attent. Demand max	0.08	0.64	Across Task Interference mean nonzero	-0.04	0.80
Visual Attent. Demand mean	0.06	0.75	W/Index max	-0.06	0.72
Auditory Attent. Demand max	0.03	0.87	Across Task Interference max	-0.08	0.65
Visual Attent. Demand zero	-0.05	0.79	Within Task Demand mean nonzero	-0.08	0.64
Cognitive Attent. Demand zero	-0.09	0.61	Cognitive Attent. Demand zero	-0.10	0.56
Within Task Demand zero	-0.09	0.61	Within Task Demand zero	-0.10	0.56
W/Index mean zero	-0.09	0.61	W/Index mean zero	-0.10	0.56
Cumulative VACP #zeros	-0.09	0.61	Cumulative VACP #zeros	-0.10	0.56
Auditory Attent. Demand mean nonzero	-0.12	0.47	Cumulative VACP mean nonzero	-0.13	0.46
Across Task Interference zero	-0.24	0.15	Physical Attent. Demand max	-0.13	0.46
Auditory Attent. Demand zero	-0.25	0.14	Visual Attent. Demand zero	-0.14	0.41
Physical Attent. Demand zero	-0.30	0.08	Cognitive Attent. Demand mean nonzero	-0.15	0.38

Observers: Combined Conflict Matrix					
Physical Effort			**Task Complexity**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Cognitive Attent. Demand mean nonzero	0.37	0.03	Cognitive Attent. Demand mean nonzero	0.68	<.0001
Cumulative VACP mean nonzero	0.36	0.03	Within Task Demand mean nonzero	0.67	<.0001
Within Task Demand mean nonzero	0.35	0.04	Cumulative VACP mean nonzero	0.67	<.0001
Across Task Interference mean nonzero	0.33	0.05	Physical Attent. Demand max	0.62	<.0001
Across Task Interference max	0.32	0.06	Across Task Interference mean nonzero	0.59	<.0001
Physical Attent. Demand max	0.31	0.07	Across Task Interference max	0.56	<.0001
W/Index mean nonzero	0.30	0.08	Physical Attent. Demand mean	0.54	<.0001
W/Index max	0.28	0.09	W/Index mean nonzero	0.54	<.0001
Auditory Attent. Demand mean	0.24	0.15	W/Index mean	0.51	<.0001
Cumulative VACP max	0.22	0.19	W/Index max	0.50	<.0001
Physical Attent. Demand mean	0.22	0.20	Across Task Interference mean	0.46	0.01
W/Index mean	0.22	0.20	Cognitive Attent. Demand mean	0.45	0.01
Visual Attent. Demand mean nonzero	0.22	0.20	Cumulative VACP max	0.45	0.01
Within Task Demand max	0.20	0.25	Auditory Attent. Demand mean	0.43	0.01
Across Task Interference mean	0.19	0.26	Visual Attent. Demand mean nonzero	0.41	0.01
Cognitive Attent. Demand max	0.18	0.29	Within Task Demand max	0.40	0.02
Cognitive Attent. Demand mean	0.16	0.34	Cognitive Attent. Demand max	0.35	0.03
Visual Attent. Demand max	0.15	0.37	Within Task Demand mean	0.35	0.04
Within Task Demand mean	0.14	0.42	Cumulative VACP mean	0.33	0.05
Cumulative VACP mean	0.13	0.45	Visual Attent. Demand max	0.27	0.11
Visual Attent. Demand zero	0.08	0.63	Visual Attent. Demand mean	0.18	0.31
Visual Attent. Demand mean	0.05	0.77	Physical Attent. Demand mean nonzero	0.11	0.51
Auditory Attent. Demand max	0.05	0.78	Auditory Attent. Demand max	0.01	0.96
Physical Attent. Demand mean nonzero	0.02	0.90	Visual Attent. Demand zero	-0.05	0.76
Cognitive Attent. Demand zero	-0.03	0.88	Cognitive Attent. Demand zero	-0.17	0.33
Within Task Demand zero	-0.03	0.88	Within Task Demand zero	-0.17	0.33
W/Index mean zero	-0.03	0.88	W/Index mean zero	-0.17	0.33
Cumulative VACP #zeros	-0.03	0.88	Cumulative VACP #zeros	-0.17	0.33
Auditory Attent. Demand mean nonzero	-0.03	0.85	Auditory Attent. Demand mean nonzero	-0.27	0.12
Auditory Attent. Demand zero	-0.16	0.36	Across Task Interference zero	-0.44	0.01
Across Task Interference zero	-0.16	0.36	Auditory Attent. Demand zero	-0.44	0.01
Physical Attent. Demand zero	-0.18	0.29	Physical Attent. Demand zero	-0.49	<.0001

Observers: Sikorsky Conflict Matrix					
Activity Level			Anxiety Level		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Within Task Demand mean nonzero	0.71	<.0001	Within Task Demand mean nonzero	0.64	<.0001
Cumulative VACP mean nonzero	0.71	<.0001	Cumulative VACP mean nonzero	0.64	<.0001
Cognitive Attent. Demand mean nonzero	0.69	<.0001	Cognitive Attent. Demand mean nonzero	0.63	<.0001
Cumulative VACP max	0.65	<.0001	Cumulative VACP max	0.59	<.0001
W/Index mean nonzero	0.63	<.0001	W/Index mean nonzero	0.58	<.0001
Across Task Interference mean nonzero	0.60	<.0001	Visual Attent. Demand max	0.56	<.0001
Visual Attent. Demand max	0.58	<.0001	Visual Attent. Demand mean nonzero	0.56	<.0001
Visual Attent. Demand mean nonzero	0.58	<.0001	Across Task Interference mean nonzero	0.55	<.0001
Within Task Demand max	0.55	<.0001	Within Task Demand max	0.52	<.0001
Across Task Interference max	0.51	<.0001	Physical Attent. Demand mean	0.51	<.0001
W/Index max	0.51	<.0001	Across Task Interference max	0.48	<.0001
Physical Attent. Demand mean	0.50	<.0001	W/Index max	0.48	<.0001
W/Index mean	0.47	<.0001	W/Index mean	0.48	<.0001
Physical Attent. Demand max	0.44	0.01	Physical Attent. Demand max	0.46	<.0001
Cognitive Attent. Demand max	0.42	0.01	Auditory Attent. Demand mean	0.42	0.01
Auditory Attent. Demand mean	0.41	0.01	Cognitive Attent. Demand max	0.42	0.01
Across Task Interference mean	0.41	0.01	Across Task Interference mean	0.42	0.01
Visual Attent. Demand mean	0.31	0.07	Visual Attent. Demand mean	0.35	0.04
Cumulative VACP mean	0.31	0.07	Cumulative VACP mean	0.35	0.04
Within Task Demand mean	0.31	0.07	Within Task Demand mean	0.35	0.04
Physical Attent. Demand mean nonzero	0.29	0.09	Cognitive Attent. Demand mean	0.28	0.09
Cognitive Attent. Demand mean	0.23	0.17	Physical Attent. Demand mean nonzero	0.24	0.15
Visual Attent. Demand zero	0.03	0.85	Auditory Attent. Demand max	0.01	0.95
Auditory Attent. Demand max	-0.03	0.85	Visual Attent. Demand zero	-0.08	0.64
Cognitive Attent. Demand zero	-0.06	0.71	Cognitive Attent. Demand zero	-0.16	0.34
Within Task Demand zero	-0.06	0.71	Within Task Demand zero	-0.16	0.34
W/Index mean zero	-0.06	0.71	W/Index mean zero	-0.16	0.34
Cumulative VACP #zeros	-0.06	0.71	Cumulative VACP #zeros	-0.16	0.34
Auditory Attent. Demand mean nonzero	-0.16	0.34	Auditory Attent. Demand mean nonzero	-0.24	0.16
Auditory Attent. Demand zero	-0.36	0.03	Auditory Attent. Demand zero	-0.40	0.01
Across Task Interference zero	-0.38	0.02	Across Task Interference zero	-0.40	0.01
Physical Attent. Demand zero	-0.44	0.01	Physical Attent. Demand zero	-0.46	<.0001

Observers: Sikorsky Conflict Matrix					
Mental Effort			**Performance**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Within Task Demand mean nonzero	0.43	0.01	Cognitive Attent. Demand mean	0.06	0.74
Cumulative VACP mean nonzero	0.43	0.01	Auditory Attent. Demand zero	0.04	0.81
Cognitive Attent. Demand mean nonzero	0.43	0.01	Physical Attent. Demand mean nonzero	0.03	0.85
Cumulative VACP max	0.39	0.02	Auditory Attent. Demand max	0.03	0.86
W/Index mean nonzero	0.38	0.02	Visual Attent. Demand mean	0.02	0.89
Physical Attent. Demand mean	0.36	0.03	Cumulative VACP mean	0.02	0.89
Visual Attent. Demand mean nonzero	0.36	0.03	Physical Attent. Demand zero	0.02	0.89
Across Task Interference mean nonzero	0.35	0.03	Auditory Attent. Demand mean nonzero	0.02	0.90
Visual Attent. Demand max	0.35	0.04	Across Task Interference mean	0.02	0.91
Within Task Demand max	0.32	0.06	Auditory Attent. Demand mean	0.02	0.92
W/Index mean	0.30	0.07	Across Task Interference zero	0.00	0.98
Physical Attent. Demand max	0.30	0.08	Within Task Demand mean	0.00	0.98
Across Task Interference max	0.29	0.09	W/Index mean	0.00	0.99
W/Index max	0.29	0.09	Cognitive Attent. Demand max	0.00	1.00
Auditory Attent. Demand mean	0.25	0.13	Physical Attent. Demand mean	-0.01	0.95
Across Task Interference mean	0.25	0.14	Physical Attent. Demand max	-0.04	0.83
Cognitive Attent. Demand max	0.24	0.16	Across Task Interference mean nonzero	-0.04	0.80
Physical Attent. Demand mean nonzero	0.24	0.16	Within Task Demand max	-0.06	0.74
Visual Attent. Demand mean	0.24	0.16	W/Index mean nonzero	-0.06	0.71
Cumulative VACP mean	0.24	0.16	Across Task Interference max	-0.08	0.65
Within Task Demand mean	0.21	0.21	W/Index max	-0.08	0.65
Cognitive Attent. Demand mean	0.16	0.35	Cognitive Attent. Demand mean nonzero	-0.08	0.64
Auditory Attent. Demand max	0.03	0.87	Visual Attent. Demand mean nonzero	-0.08	0.62
Visual Attent. Demand zero	-0.05	0.79	Cognitive Attent. Demand zero	-0.10	0.56
Cognitive Attent. Demand zero	-0.09	0.61	Within Task Demand zero	-0.10	0.56
Within Task Demand zero	-0.09	0.61	W/Index mean zero	-0.10	0.56
W/Index mean zero	-0.09	0.61	Cumulative VACP #zeros	-0.10	0.56
Cumulative VACP #zeros	-0.09	0.61	Visual Attent. Demand max	-0.12	0.49
Auditory Attent. Demand mean nonzero	-0.12	0.47	Within Task Demand mean nonzero	-0.13	0.46
Across Task Interference zero	-0.24	0.15	Cumulative VACP mean nonzero	-0.13	0.46
Auditory Attent. Demand zero	-0.25	0.14	Visual Attent. Demand zero	-0.14	0.41
Physical Attent. Demand zero	-0.30	0.08	Cumulative VACP max	-0.15	0.38

Observers: Sikorsky Conflict Matrix					
Physical Effort			**Task Complexity**		
Analytical Measure	ρ	α	**Analytical Measure**	ρ	α
Within Task Demand mean nonzero	0.39	0.02	Within Task Demand mean nonzero	0.70	<.0001
Cumulative VACP mean nonzero	0.39	0.02	Cumulative VACP mean nonzero	0.70	<.0001
Cognitive Attent. Demand mean nonzero	0.35	0.04	Cognitive Attent. Demand mean nonzero	0.67	<.0001
Cumulative VACP max	0.35	0.04	Cumulative VACP max	0.66	<.0001
W/Index mean nonzero	0.33	0.05	W/Index mean nonzero	0.64	<.0001
Across Task Interference mean nonzero	0.33	0.05	Visual Attent. Demand mean nonzero	0.61	<.0001
Across Task Interference max	0.32	0.06	Visual Attent. Demand max	0.60	<.0001
W/Index max	0.32	0.06	Across Task Interference mean nonzero	0.59	<.0001
Within Task Demand max	0.31	0.07	Within Task Demand max	0.58	<.0001
Visual Attent. Demand max	0.31	0.07	Across Task Interference max	0.56	<.0001
Visual Attent. Demand mean nonzero	0.28	0.10	W/Index max	0.56	<.0001
Auditory Attent. Demand mean	0.24	0.15	Physical Attent. Demand mean	0.53	<.0001
Cognitive Attent. Demand max	0.23	0.17	W/Index mean	0.51	<.0001
W/Index mean	0.22	0.20	Physical Attent. Demand max	0.47	<.0001
Across Task Interference mean	0.19	0.26	Cognitive Attent. Demand max	0.47	<.0001
Physical Attent. Demand mean	0.19	0.27	Across Task Interference mean	0.46	0.01
Physical Attent. Demand max	0.17	0.33	Auditory Attent. Demand mean	0.43	0.01
Within Task Demand mean	0.15	0.39	Within Task Demand mean	0.40	0.02
Cognitive Attent. Demand mean	0.13	0.45	Visual Attent. Demand mean	0.38	0.02
Visual Attent. Demand mean	0.12	0.50	Cumulative VACP mean	0.38	0.02
Cumulative VACP mean	0.12	0.50	Cognitive Attent. Demand mean	0.33	0.05
Physical Attent. Demand mean nonzero	0.09	0.59	Physical Attent. Demand mean nonzero	0.21	0.21
Visual Attent. Demand zero	0.08	0.63	Auditory Attent. Demand max	0.01	0.96
Auditory Attent. Demand max	0.05	0.78	Visual Attent. Demand zero	-0.05	0.76
Cognitive Attent. Demand zero	-0.03	0.88	Cognitive Attent. Demand zero	-0.17	0.33
Within Task Demand zero	-0.03	0.88	Within Task Demand zero	-0.17	0.33
W/Index mean zero	-0.03	0.88	W/Index mean zero	-0.17	0.33
Cumulative VACP #zeros	-0.03	0.88	Cumulative VACP #zeros	-0.17	0.33
Auditory Attent. Demand mean nonzero	-0.03	0.85	Auditory Attent. Demand mean nonzero	-0.27	0.12
Auditory Attent. Demand zero	-0.16	0.36	Across Task Interference zero	-0.44	0.01
Across Task Interference zero	-0.16	0.36	Auditory Attent. Demand zero	-0.44	0.01
Physical Attent. Demand zero	-0.18	0.29	Physical Attent. Demand zero	-0.49	<.0001

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Establishment sponsoring a contractor's report, or tasking agency, are entered in section 12.)
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West
North York, Ontario

2. DOCUMENT SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable)
Unclassified

3. DOCUMENT TITLE (the complete document title as indicated on the title page. Its classification should be indicated be the appropriate abbreviation (S,C,R or U) in parentheses after the title.)
An evaluation of workload model predictions in a helicopter environment

4. DESCRIPTIVE NOTES (the category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)
Technical Report

5. AUTHOR(S) (Last name, first name, middle initial. If military, show rank, e.g. Burns, Maj. Frank E.)
Cain, Brad

6. DOCUMENT DATE (month and year of publication of document)	7.a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)	7.b. NO. OF REFS. (total cited in document)
Oct 97	64	21

8.a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	8.b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)

9.a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	9.b. OTHER DOCUMENT NO.(S) (any other numbers which may be assigned this document either by the originator or by the sponsor.)
97-12-66	

10. DOCUMENT AVAILABILITY (any limitation on further dissemination of the document, other than those imposed by security classification)

- [√] Unlimited distribution
- [ ] Distribution limited to defence departments and defence contractors; further distribution only as approved
- [ ] Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
- [ ] Distribution limited to government departments and agencies; further distribution only as approved
- [ ] Distribution limited to defence departments; further distribution only as approved
- [ ] Other

11. ANNOUNCEMENT AVAILABILITY (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (10.) However, where further distribution (beyond the audience specified in 10) is possible, a wider announcement audience may be selected.)

12. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

This report compares the empirical workload results obtained during a field exercise involving four CH-136 Kiowa crews with results predicted from a task network simulation of the exercise.

None of the correlation coefficients is outstanding and only a few of the analytical measures explained greater than 50% of the variance in the empirical workload scores. The correlations between the pilots' empirical workload scores and the analytical workload values were generally greater than that found for the observers, possibly a result of the smaller number of subjects in the observer group or perhaps reflecting greater attention focused on the workload of the pilot by the modelling community. The variation of the workload measures within each flight was substantial, suggesting individual differences between subjects as well as differences in the details of each mission played significant roles in determining the perception of workload rated by the subjects.

Of the overall workload measures, the simpler measures were found to capture the greatest portion of the empirical workload variance although these measures provide little detail in what is actually causing the overload and at best only capture 50% of the workload variance. While the multi-dimensional workload measures may provide greater detail about what is causing high workloads, they do not seem to be capturing a great deal of the workload variance to begin with.

Although the correlations found in this study are low and only half the workload variance was captured, the models may still be useful. The methods currently in use are, however, still immature and require that the user have detailed knowledge of their capabilities and limitations. Considerably more research followed by thoughtful development is necessary before this analysis approach should be released as a robust and proven tool to the design community.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible, keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

task network, cognitive workload, mental workload